

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:23:46 ; Search time 53.52 seconds
(without alignments)
566.576 Million cell updates/sec

Title: US-09-673-763-14

Perfect score: 1333

Sequence: 1 MTTPTLIIVPPSPAPSYSA.....QIRETLSPKSPSTKSS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1333	100.0	273	20	AA1980.DAT.*
2	840	63.0	195	20	AA1981.DAT.*
3	317	23.8	81	20	AA1982.DAT.*
4	162.5	12.2	355	20	AA1983.DAT.*
5	154.5	11.6	397	20	AA1984.DAT.*
6	124.5	9.3	2482	16	AA1985.DAT.*
7	124.5	9.3	2482	16	AA1986.DAT.*
8	124.5	9.3	3248	17	AA1987.DAT.*
9	120	9.0	1931	22	AA1988.DAT.*
10	116	8.7	875	22	AA1989.DAT.*
11	116	8.7	878	22	AA1990.DAT.*

12	116	8.7	1017	22	AAE02246	Domestic mite Btll
13	116	8.7	1177	22	AA196721	Putative p. abysal
14	115.5	8.7	1372	19	AAW56473	Protein with Rho p
15	115	8.6	1325	18	AAW19540	Male-enhanced anti
16	115	8.6	1325	20	AAW94391	Mouse male enhance
17	113.5	8.5	881	22	ABG05280	Novel human diagno
18	113.5	8.5	881	22	ABG05280	Novel human diagno
19	112.5	8.4	622	22	ABG62816	Drosophila melanog
20	112.5	8.4	789	22	AA195460	Human protein sequ
21	111.5	8.4	672	21	AA195460	Human protein sequ
22	111.5	8.4	1192	22	AAU35310	Enterococcus faeca
23	111.5	8.4	1388	19	AAW56475	Protein with Rho p
24	111	8.3	334	15	AA195475	Membrane antigen p
25	111	8.3	433	19	AA195475	Human glial fibril
26	109.5	8.2	398	21	AA195475	Partial protein en
27	109	8.2	2415	22	ABG20279	Novel human diagno
28	108.5	8.1	513	22	ABG08133	Novel human diagno
29	108.5	8.1	870	22	ABG08133	Drosophila melanog
30	108.5	8.1	1081	21	AA195490	Arabidopsis thalia
31	108.5	8.1	1197	21	AA195490	Arabidopsis thalia
32	108.5	8.1	1690	22	ABG61144	Drosophila melanog
33	108.5	8.1	1690	22	ABG61144	Drosophila melanog
34	108.5	8.1	2013	22	ABG61173	Drosophila melanog
35	108	8.1	430	21	AA195490	Arabidopsis thalia
36	108	8.1	458	21	AA195490	Arabidopsis thalia
37	108	8.1	473	21	AA195490	Arabidopsis thalia
38	108	8.1	776	22	ABG05279	Novel human diagno
39	108	8.1	777	22	ABG05279	Novel human diagno
40	107.5	8.1	433	21	AA195490	Breast and ovarian
41	107.5	8.1	469	18	AAW23820	Amino acid sequenc
42	107.5	8.1	469	21	AA195490	Human keratin KERT
43	107.5	8.1	546	21	AA195490	Receptor for hyalu
44	107.5	8.1	606	17	AA195490	RHAMM 1-2a isoform
45	107.5	8.1	631	17	AA195490	

ALIGNMENTS

```
RESULT 1
AA1980.DAT.*
ID AA1980.DAT.* standard; Protein: 273 AA.
AC AA1980.DAT.*
XX
XX
XX 01-FEB-2000 (first entry)
XX
XX Chlamydia trachomatis infection-specific protein Inca.
XX Inca; infection; vaccine; therapy; diagnosis.
XX Chlamydia trachomatis.
XX
XX WO9953948-A1.
XX
XX 28-OCT-1999.
XX
XX 20-APR-1999; 99WO-US08744.
XX
XX 20-APR-1998; 98US-0082438.
XX 21-APR-1998; 98US-0082588.
XX 22-MAY-1998; 98US-0086450.
XX (UYOR-) UNIV OREGON STATE.
XX
XX Rocky DD, Bannantine JP;
XX
XX WPT; 1999-633904/54.
XX N-PSDB; AA234590.
XX
XX Novel bacterial infection specific proteins for treating and diagnosing
XX chlamydial infections
```

PS Claim 5; Page 45-46; 56pp; English.

XX This sequence represents novel infection-specific protein Inca of

CC Chlamydia trachomatis LGV-434 (serotype L2). Inca is found in the

CC inclusion membrane of infected cells. It is associated primarily

CC with the vegetative reticulate body form of Chlamydia rather than

CC with the refractile elementary body form. The invention includes: a

CC vaccine directed against the reticulate body form of Chlamydia

CC comprising 1 or more infection-specific proteins (see AAY32170-78),

CC including Inca, IncB and IncC; methods of using and producing such

CC a vaccine; methods for detection of infection-specific antibodies

CC or antigens in a biological specimen; and a method of using

CC therapeutic agents specifically directed against infection-specific

CC peptides, or the genes that code for such peptides, to treat

CC chlamydial infection.

XX

SQ Sequence 273 AA;

Query Match 100.0%; Score 1333; DB 20; Length 273;

Best Local Similarity 100.0%; Pred. No. 2.6e-106; Mismatches 0; Gaps 0;

Matches 273; Conservative 0; Indels 0;

QY 1 MTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAIASLILIGTIGFLALHGLVGF 60

DB 1 mtptlivippapppapsanrvppslmdkikkiaiasliligtigflalhglvfl 60

QY 61 IAPQITIVLLALFITSAGNALYLOKTANLHLYQDLOREVGSKEINFMLSVLQKEFLHL 120

DB 61 iapqitivilalfitslagnalyloktanlhlyqdlqrevgskeinfmlsvlqkeflhl 120

QY 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLDEYKNSTEEMRKLFQSIADLKG 180

DB 121 skfattskdlsavsqdfysclgfrdnykgfeslldeyknsteemrklfesqeiadlkg 180

QY 181 SVASLREEIRFLTPLAEVRRLAHQESLTAATEELKTIRDSLRDEIGQLSOLSKTLTSQ 240

DB 181 svaslreeirfltplaevevrrlahnqesltaateelktirdslrdeigqslsktltsq 240

QY 241 IALQKESDLCQSIREFTLSSPRKSASPSTKSS 273

DB 241 ialqrkessdlcsqiretlssprksaspstks 273

RESULT 2

AAY37087

AAY37087 standard; Protein; 195 AA.

AAY37087;

XX 07-OCT-1999 (first entry)

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;

KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX

PI Griffiths R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 892-893; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perinephritis, bartholinitis; pneumopathy in breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX

SQ Sequence 195 AA;

Query Match 63.0%; Score 840; DB 20; Length 195;

Best Local Similarity 91.5%; Pred. No. 3.2e-64;

Matches 173; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 64 QITIVLLALFITSAGNALYLOKTANLHLYQDLOREVGSKEINFMLSVLQKEFLHLSKE 123

DB 4 kftnvllaifitslagnalyloktanlhlyqdlqrevgskeinfmlsvlqkeflhlske 63

QY 124 FATTSKDLASVQDFYSCLOGFRDNYKGFESLDEYKNSTEEMRKLFQSIADLKGSA 183

DB 64 fattskdlsavsqdfysclggrdnykgfeslldeyknsteemrklfesqeiadlkgsva 123

QY 184 SLREEIRFLTPLAEVRRLAHQESLTAATEELKTIRDSLRDEIGQLSOLSKTLTSOIAL 243

DB 124 slreeirfltplaevevrrlahnqesltaateelktirdslrdeigqslsktltsaklh 183

QY 244 QKESDLC 252

DB 184 ynekragic 192

RESULT 3

AAY37086

ID AAY37086 standard; Protein; 81 AA.

XX

AC AAY37086;

XX 07-OCT-1999 (first entry)

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;

KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX

PI Griffais R;
 XX WPI; 1999-371125/31.
 XX Genome sequence of Chlamydia trachomatis
 XX Disclosure; Page 892; 1755pp; English.
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma, genital
 CC paratrachoma, and inclusion conjunctivitis; peritonitis, salpingitis,
 CC nongonococcal urethritis, epididymitis, cervicitis, perinephritis,
 CC perihepatitis, Bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX Sequence 81 AA;
 SQ
 Query Match 23.8%; Score 317; DB 20; Length 81;
 Best Local Similarity 98.5%; Pred. No. 7.2e-20;
 Matches 64; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MTTPLIIVPPSPAPSANRVPOPSLMDKIKKIAATASLILIGTIGFLALLGHLVGF 60
 DB 16 mttptllivtppsppspsanrvpqpsslmddkikkiaaaslllgtigfllglnlvgl 75
 QY 61 IAPQI 65
 DB 76 iapqi 80
 RESULT 4
 AAY32173
 ID AAY32173 standard; Protein; 355 AA.
 XX
 AC AAY32173;
 XX
 DT 01-FEB-2000 (first entry)
 DE Chlamydia psittaci infection-specific protein Inca.
 DE Inca; infection; vaccine; therapy; diagnosis.
 KW Chlamydia psittaci.
 XX
 OS WO9953948-A1.
 PD 28-OCT-1999.
 XX
 PF 20-APR-1999; 99WO-US08744.
 XX
 PR 20-APR-1998; 98US-0082438.
 PR 21-APR-1998; 98US-0082588.
 PR 22-MAY-1998; 98US-0086450.
 XX
 PA (UYOR-) UNIV OREGON STATE.
 XX
 XX Rocky DD, Bannantine JP;
 XX WPI; 1999-633904/54.
 DR N-PSDB; AAZ34587.
 XX
 XX Novel bacterial infection specific proteins for treating and diagnosing
 PT chlamydial infections
 XX
 PS Claim 5; Page 39-40; 56pp; English.
 XX
 CC This sequence represents novel infection-specific protein Inca of

CC Chlamydia psittaci strain GPIC. Inca is found in the inclusion
 CC membrane of infected cells. It is associated primarily with the
 CC vegetative reticulate body form of Chlamydia rather than with the
 CC refractile elementary body form. The invention includes: a vaccine
 CC directed against the reticulate body form of Chlamydia comprising 1
 CC or more infection-specific proteins (see AAY32170-78), including
 CC Inca, IncB and IncC; methods of using and producing such a vaccine;
 CC methods for detection of infection-specific antibodies or antigens
 CC in a biological specimen; and a method of using therapeutic agents
 CC specifically directed against infection-specific peptides, or the
 CC genes that code for such peptides, to treat chlamydial infection.
 XX Sequence 355 AA;
 SQ
 Query Match 12.2%; Score 162.5; DB 20; Length 355;
 Best Local Similarity 21.4%; Pred. No. 9.3e-06;
 Matches 63; Conservative 61; Mismatches 111; Indels 59; Gaps 8;
 QY 8 VIPSPAPSPAPSANRV-----POPSLMDKIKKIAATASLILIGTIGFLALLGHLVG 58
 DB 32 lipisieaptssaaavgaaktaiepgsrpllqricyivkkliaaiaifvfgaaivclvlg 91
 QY 59 FLIAPQITIVLLALFITSLSA-----GNALYLQKNTANLHLYQDLQR-----EV 100
 DB 92 svistpslilmlaamlvsvivitaigrdtpsqvvr-----hmkqqiqqfgeentrlhtav 147
 QY 101 GSKKEINFMLS--VLOKEFLH-----LSKEFATSKDLSAVSQDFYSCLOGFRDNYKGF 153
 DB 148 enlkavnvelseqinqlkqhtrlsdfgrleantgofatladfqlsleefksvgtkve 207
 QY 154 SLDEYKNSTEEMKLFQSEIIADLKGVSASLREIEIRFLPLAEVRRRLAHNQESLTAI 213
 DB 208 tmlspfeklaqsketfsgavqammsvvtelrt-----hlnakell 250
 QY 214 EELKTRDSLRDEIGLSQLSKTLTSQLQRLQKESDLCQIRETLSSPRKAS 267
 DB 251 tenktvleqlkad----agireeqvrflekkrkdeleacstlahsiatiqestt 300
 RESULT 5
 AAY34783
 ID AAY34783 standard; Protein; 397 AA.
 XX
 AC AAY34783;
 XX
 DT 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae transmembrane protein sequence.
 DE Chlamydia pneumoniae.
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX Chlamydia pneumoniae.
 XX WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 XX (GEST) GENSET.
 PA Griffais R;
 XX
 XX WPI; 1999-357842/30.
 XX
 XX Genome sequence of Chlamydia pneumoniae

```

PS Page 760-761; Disclosure: 1912bp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY34584) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 397 AA;
Query Match 11.6%; Score 154.5; DB 20; Length 397;
Best Local Similarity 20.8%; Pred. No. 5.2e-05;
Matches 75; Conservative 59; Mismatches 123; Indels 103; Gaps 12;
QY 3 TPTLVIP-PPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLALLGHLVGF-- 59
DB 15 tpsapnipaptptgipthk-prssf--iekviavakylfaiaatsgalgtlglsg 70
QY 60 LTAPQITIVLLALFTTSLAGNALYLQKNTANLHLQDLOREYVGLSKNEFMLSVLQKEFLH 119
DB 71 altpgigiallvifvsmvllgllkdsiggeerrirreevrsftsenqrltv----- 123
QY 120 LSKEFATTSKDL SAVSQDFYSCLOQFRD---NYK----- 150
DB 124 ittletevkdkaakdqltleieafnengnlkttaedleeqvsklseqlalerlnql 183
QY 151 -----GPES-LDDEYKNSTEEMRKLFQSOEIIADLKGVSASLREEI 189
DB 184 iqanagdaqeisellkklisgwskvveqintsqalkvllgqevvqeqthvkaqqeqi 243
QY 190 RFL-----TPLAEVRLAHNQESLTAAIEELKTRD-----SLRDEIGOLS 231
DB 244 qalqaeilgmhngstalgksvenllvqdaqaltrvvvgellessenklisqacsalrqeiekla 303
QY 232 QLSKLTTSQI-----ALQR-----KESDLCQIRFELSSPRKSASPST 270
DB 304 qhetslqqrldamlageqnlacvtaekmkqeqadkaesefiacvdrtrfgretppptt 363
RESULT 6
XX AAR72826 standard; Protein; 2482 AA.
XX AAR72826;
XX 27-FEB-1996 (first entry)
XX Human mitosis.
XX Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;
XX inhibition.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Region 1480..1659
XX /label= internal_repeat
XX Region 1660..1839
XX /label= internal_repeat
XX W09511309-A2.
XX 27-APR-1995.
XX 24-OCT-1994; 94WO-US12162.
XX

```

```

PR 22-OCT-1993; 93US-0141239.
XX (TEXA ) UNIV TEXAS SYSTEM.
XX Lee W, Zhu X;
XX WPI: 1995-170229/22.
XX N-PSDB; AAG86851.
XX Purified mammalian protein mitosis and agents that bind it and
XX inhibit its action -- used to promote cell growth or to inhibit cell
XX division and/or proliferation
XX Claim 4; Fig 8B; 6lpp; English.
XX AAR72829 is human mitosis. Mitosis is involved in the regulation of
XX the mammalian mitotic cell cycle. Mitosis as with E2F-1 (see AAR72824)
XX interacts with the retinoblastoma protein (the retinoblastoma tumour
XX suppressor gene product). Mitosis is first synthesised at the G1/S
XX boundary, it is then phosphorylated from S through M phase, and during
XX mitosis, is closely associated with the centromeres/kinetochores at the
XX mitotic spindle poles. Mitosis is necessary for a eukaryotic cell to
XX enter the M phase of the mitotic cell cycle and its degradation is
XX necessary for a cell to advance on to the next stage. Mitosis is thus
XX useful for controlling cell growth as overexpression of mitosis prevents
XX a cell from exiting the M phase.
XX An anti-mitosis antibody, antibody fragment or a phosphorylated mitosis
XX mutin ( or nucleic acid encoding it) can also be used to inhibit cell
XX division which is particularly useful for the study of the cell cycle.
XX A further use is to control hyperproliferative cells, and so control
XX diseases such as psoriasis and breast cancer. It can also be used to
XX block gametogenesis of an immature gamete.
XX
SQ Sequence 2482 AA;
Query Match 9.3%; Score 124.5; DB 16; Length 2482;
Best Local Similarity 26.2%; Pred. No. 0.23;
Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
QY 77 LAGNALYLQK-----TANLHLQDLOREYVGLSKNEFMLSVLQKEFLHLSKEFAT--- 126
DB 314 llsetlslekkemsiislnkreleltqengtlkeinasinqemlnliqksesfanyid 373
QY 127 -TSKDLSAVS---QDFYSCLOQFRDNYKGFESLDEYKNSTEEMRKL-----FSQEI 176
DB 374 ereksiselsdqykqekilllqrceetgnayedlsqykaaeknksklecllnetsice 433
QY 177 DLKGSVASLREEI-----RFLTPLAEEVRLAHNQESLTAAIEELKTRDLSLREI---- 227
DB 434 nrkneleqlkeafakehqefitkfaaeer---ngnlml-----eletvqqalrsemtdnq 486
QY 228 -----GOLSQSLSKTLTSQIALQKESSDLCQ 254
DB 487 nnskseagglkqeqintllkeeqnkmqkevndllqe 520
RESULT 7
XX AAW23996
XX ID AAW23996 standard; Protein; 2482 AA.
XX AC AAW23996;
XX 28-MAY-1998 (first entry)
XX Human mitosis amino acid sequence.
XX Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue;
XX inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
XX Leukaemia; lymphoma; chromosome segregation.
XX Homo sapiens.
XX

```


FH Key Location/Qualifiers
 FT 258..280
 FT /note= "leucine heptad repeat"
 FT 340..362
 FT 564..593
 FT 1387..1443
 FT 1885..1962
 FT 2146..2188
 FT 2165..2187
 FT /note= "leucine heptad repeat"
 FT Misc-difference 2188
 FT Misc-difference 2300
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally C or G"
 FT Misc-difference 2189
 FT Misc-difference 2301
 FT Misc-difference 2303
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally A or T"
 XX US5710022-A.
 DN 20-JAN-1998.
 XX 24-OCT-1994; 94US-0328254.
 XX 24-OCT-1994; 94US-0328254.
 PR 22-OCT-1993; 93US-0141239.
 XX (TEXA) UNIV TEXAS SYSTEM.
 PA Lee W, Zhu X;
 PI WPI; 1998-109817/10.
 DR N-PSDB; AAV09076.
 XX New isolated mitosis protein and gene - useful for, e.g. developing
 PT products for therapy and diagnosis of hyper-proliferative disorders
 PT such as cancers or psoriasis
 XX Claim 1; Column 40-52; 43pp; English.
 CC This is the amino acid sequence for mitosis, a phosphoprotein
 CC necessary for the cell to enter mitosis. The protein's degradation is
 CC also necessary for the cell to advance into the next stages of mitosis.
 CC The mitosis protein, can be used to control the growth of cells. An
 CC anti-mitosis antibody, a mutant or a non-functional analogue of mitosis
 CC can inhibit the mitotic cell cycle by preventing the cells from entering
 CC the M phase, and over expression of mitosis or its functional
 CC equivalent, would inhibit the cycle by preventing cells from leaving the
 CC M phase. Antagonists to this protein can be used to control
 CC hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
 CC psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 CC cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
 CC lung cancer and various leukemias and lymphomas). Reintroduction or
 CC supplementation of lost mitosis function by introduction of the protein
 CC or nucleic acid encoding the protein into a cell can restore defective
 CC chromosome segregation, which is a marker of progressing malignancy.
 CC Malignant proliferation of cells can then be halted. The protein
 CC can also be used for the detection and diagnosis of hyperproliferative
 CC cells.
 XX Sequence 2482 AA;

Query Match 9.38; Score 124.5; DB 19; Length 2482;
 Best Local Similarity 26.28; Pred. No. 0.23;
 Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;

OY 77 LAGNALYLOK-----TANLHLVQDLOREVGLSKLEINFMLSVLQKFLHLSKEFAT--- 126
 DB 314 llsletislekemassifslslnkreieeltqngtltkeinaslnqemlnliqsfsfanyid 373

OY 127 -TSKDL SAVS-----QDFYSCLOQFRDNYKGFESLLDEYKNSTEEMRKL-----FSOEIIA 176
 DB 374 ereksiselsdqykqeklllllqrceetgnayedlsqkykaaqeksklecllnectslce 433
 OY 177 DLKGSVASLREEI-----RFLPLAEVRRLAHNQESLTAATIEELKTIKIRDSLRDEI----- 227
 DB 434 nrkneleqkkaefakehgefllkfaeer---ngnlml-----eletvqgarsemtdng 486
 OY 228 -----GQLSQLSKTLTSQIALQRKESSDLCQ 254
 DB 487 nnskseagglkqeimtlkeeqnkmqkevndllqe 520
 RESULT 8
 AAR99795
 ID AAR99795 standard; Protein; 3248 AA.
 XX AAR99795;
 XX 08-OCT-1996 (first entry)
 XX Kinetochores protein CENP-F.
 XX Kinetochores protein; CENP-F; cell cycle; cancer; diagnosis;
 KW autoimmune antibody.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Domain 1..200
 FT /label= Extended_coiled_structure
 FT 280..1350
 FT /label= Extended_coiled_structure
 FT 1380..1610
 FT /label= Globular_domain
 FT /note= "globular domain consists of 2 direct
 FT repeats of 95 amino acids"
 FT Domain 1620..1750
 FT /label= Extended_coiled_structure
 FT 1850..2990
 FT /label= Extended_coiled_structure
 FT 3048..3248
 FT /label= C-terminal_domain
 FT /note= "the C-terminal domain is predicted to
 FT form a proline-rich (10.6%) highly
 FT basic (pI 10) globular domain"
 XX WO9617867-A1.
 XX 13-JUN-1996.
 XX 08-DEC-1995; 95WO-US16216.
 XX 09-DEC-1994; 94US-0353700.
 XX (FOXC-) FOX CHASE CANCER CENT.
 XX (UYTE-) UNIV TECHNOLOGIES INT INC.
 XX Rattner JB, Yen TJ;
 XX WPI; 1996-287116/29.
 XX N-PSDB; AAT34578.
 XX DNA encoding kinetochores protein - used as a marker for the G2 and M
 XX phases of a cell cycle, partic. for detection of malignant diseases
 PS Claim 12; Page 41-54; 72pp; English.
 CC A 372 kDa human kinetochores protein, CENP-F (AAR99795), is detected
 CC by immunofluorescence microscopy only during the G2 and M phases
 CC of a cell cycle. It is the product of a cDNA clone (AAR34578)
 CC isolated from a breast carcinoma cDNA library. Recombinant CENP-F
 CC can be produced by expression in prokaryotic or eukaryotic host

10-OCT-2000; 2000WO-AU01227.
 26-OCT-1999; 99SG-0005313.
 18-JUL-2000; 2000AU-0008842.
 18-JUL-2000; 2000AU-0008844.
 18-JUL-2000; 2000AU-0008845.
 (UYSI-) UNIV SINGAPORE NAT.
 Chua KY, Cheong N, Lee BW;
 WPI; 2001-308609/32.
 N-PSDB; AAD06245.
 Novel immunogenic protein derived from house mite, *Blomia tropicalis*
 useful for treating and diagnosing conditions involving induction of
 immuneresponse to mite, such as allergic asthma, atopic dermatitis,
 rhinitis -
 Claim 6; Fig 7; 230pp; English.
 The present invention relates to immunogenic proteins, referred as Bt
 allergen, is derived from domestic mite, *Blomia tropicalis*. The specific
 Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
 immunogenic protein is useful for preventing, reducing or ameliorating
Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
 immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
 asthma and for modulating an immune response directed to Bt allergen in
 a subject. The Bt allergens are also useful for detecting antibody
 directed to all or a part of Bt allergen in a biological sample from a
 subject. Antibodies to Bt allergens are also used as therapeutic or
 diagnostic agents, to screen Bt immunoassays and as antagonists to
 inhibit Bt activity under circumstances where temporary hypersensitivity
 inhibition is required. The present sequence is a protein encoded
 by Bt11 polymorphic variant.
 Sequence 1017 AA;
 Query Match 8.7%; Score 116; DB 22; Length 1017;
 Best Local Similarity 26.9%; Pred. No. 0.37; Indels 22; Gaps 5;
 Matches 46; Conservative 31; Mismatches 72;
 QY 94 QDLQREVGSLKEINFMVLSVLOKEFLHLSKEFATTSKDSAVSQDFYSCIQGFRDNYKGF 153
 453 qalekvrsglekinidkksleevsmllqeq---tqkdlrvkiad---lqklqhey---e 502
 154 SLDEYKNSTEMRKLFQSEIADLKGSVASLREERFRLTPLAEVRRRLAHNQESLTAI 213
 503 klrdqkealarenkkiaad-----dlaeaaksgindahrrihegeieikrlenerelaay 557
 QY 214 EELKTIKRLSRDEIGQLSQSLKTLTQIALQKESDLCSSQIRETLSSPRK 264
 558 keaetirk-----qeeaknqrtaelqatrhdyekrlaqkeeelealrk 601
 RESULT 13
 AAB96721
 ID AAB96721 standard; Protein; 1177 AA.
 XX
 AC AAB96721;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Putative P. abyssi ATPase involved in DNA repair #3.
 XX
 KW Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 XX
 PD 27-OCT-2000.
 21-APR-1999; 99FR-0005034.
 21-APR-1999; 99FR-0005034.
 (CNRS) CNRS CENT NAT RECH SCI.
 (IFRE-) IFREMER INST FR RECH EXPL MER.
 Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 Querellou J, Weissenbach J, Saurin W, Heilig R;
 WPI; 2001-126236/14.
 New nucleotide sequences isolated from *Pyrococcus abyssi* encode
 proteins useful in industry -
 Claim 7; Pages 1483-1487; 1657pp; French.
 The present invention relates to the genomic sequence of *Pyrococcus*
 abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is
 a hyperthermophilic archaeon, which is isolated from deep-sea
 hydrothermal vents. The present sequence is one such P. abyssi protein.
 The proteins of the present invention have various potential industrial
 uses, since the proteins are stable at very high temperatures, some up to
 110 degrees centigrade.
 Note: This patent is in the same patent family as WO2000065062, which
 contains additional sequences as shown in AAB99132-AAB99143,
 AAB75903-AAB75920 and AAG66436.
 Sequence 1177 AA;
 Query Match 8.7%; Score 116; DB 22; Length 1177;
 Best Local Similarity 19.7%; Pred. No. 0.45; Indels 82; Gaps 9;
 Matches 48; Conservative 42; Mismatches 72;
 QY 82 LYLOKANTANLHLYQDLQREVGSLK-----EINFMVLSVLOKEFLHLSKEFATTSKD 130
 673 lklrkea-----leaeinslkvelglenqgfelrkmseiekeitlrlrdieklise 725
 QY 131 LSASQDFYSCIQGFRDNYKGFESL-----LDEYKNSTEMRKLFQSE- 172
 726 erliikse-----ledsqgideidriihekkgelakrlgrkierlekrklkalenp 778
 QY 173 -----ELIADLKGSVASLREI-----RFLTP-----LAEEVR----- 200
 779 earevtekirovegeigklreelsrvesrleslnsrneeliprkasleeelegivnkin 838
 QY 201 -----RLAHNQE---SLTAIEELKTIKRLSRDEIGQLSQSLKTLTQIALQKESDLCSS 253
 839 alkaniaeneevlkgikgleelkakeesvhsikseyrrkrreelekeirelkekeelsk 898
 QY 254 QIRE 257
 899 rmqge 902
 RESULT 14
 AAW56473
 ID AAW56473 standard; Protein; 1372 AA.
 XX
 AC AAW56473;
 XX
 DT 14-AUG-1998 (first entry)
 XX
 DE Protein with Rho protein-combining and kinase activity.
 XX
 KW Rho protein-binding activity; protein kinase activity; inhibitor;
 KW smooth muscle fibre formation; smooth muscle contraction;
 KW circulatory disease; treatment; tumour formation; metastasis inhibitor;
 KW autoimmune disease; platelet aggregation inhibitor.
 XX
 OS Bos sp.

Db 314 LLSETLSLEKKEWSSIIISLNKREIEBELTQENGTLKEINASLNQKMNLIQKSESFANYID 373
Qy 127 -TSKDLASVS-----ODFYSCLOGFRDNYKGFESLLDDEYKKNSTEEMRKL-----FSQEIIA 176
Db 374 EREKSISLSQDYQKQKLLILORCEETGNAYEDLSQYKAAQKNSKLECLLNECTSLCE 433
Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 434 NRKNELEQLKEAFKHOEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 486
Qy 228 -----GOLSQSKTLTSQIALQKESSDLCQ 254
Db 487 NNSKSEAGGLKQEIIMTLKEQNKMQKEVNDLLQ 520
RESULT 2
PCT-US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: HUMAN
ORGANISM: HUMAN
US-08-353-700-1
Query Match 9.3%; Score 124.5; DB 1; Length 3248;
Best Local Similarity 26.2%; Pred. No. 0.034;
Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
Qy 77 LAGNALYLQK-----TANLHYODLOREVGSLSKEINFMLSVLQKFEFLHLSKEFAT--- 126
Db 946 LLSETLSLEKKEWSSIIISLNKREIEBELTQENGTLKEINASLNQKMNLIQKSESFANYID 1005
Qy 127 -TSKDLASVS-----ODFYSCLOGFRDNYKGFESLLDDEYKKNSTEEMRKL-----FSQEIIA 176
Db 1006 EREKSISLSQDYQKQKLLILORCEETGNAYEDLSQYKAAQKNSKLECLLNECTSLCE 1065

Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 1066 NRKNELEQLKEAFKHOEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 1118
Qy 228 -----GOLSQSKTLTSQIALQKESSDLCQ 254
Db 1119 NNSKSEAGGLKQEIIMTLKEQNKMQKEVNDLLQ 1152
RESULT 3
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1
Query Match 9.3%; Score 124.5; DB 5; Length 3248;
Best Local Similarity 26.2%; Pred. No. 0.034;
Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
Qy 77 LAGNALYLQK-----TANLHYODLOREVGSLSKEINFMLSVLQKFEFLHLSKEFAT--- 126
Db 946 LLSETLSLEKKEWSSIIISLNKREIEBELTQENGTLKEINASLNQKMNLIQKSESFANYID 1005
Qy 127 -TSKDLASVS-----ODFYSCLOGFRDNYKGFESLLDDEYKKNSTEEMRKL-----FSQEIIA 176
Db 1006 EREKSISLSQDYQKQKLLILORCEETGNAYEDLSQYKAAQKNSKLECLLNECTSLCE 1065
Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 1066 NRKNELEQLKEAFKHOEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 1118
Qy 228 -----GOLSQSKTLTSQIALQKESSDLCQ 254

us-09-673-763-14.ra

Tue Aug 13 14:23:40 2002

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/685,576
 FILING DATE: 24-JUL-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 7-325129
 FILING DATE: 20-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-17150
 FILING DATE: 05-JAN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-131206
 FILING DATE: 26-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16887/843
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1388 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-685-576-4

Query Match 8.4%; Score 111.5; DB 2; Length 1388;
 Best Local Similarity 22.1%; Pred. No. 0.15;
 Matches 43; Conservative 40; Mismatches 91; Indels 21; Gaps 4;
 QY 84 LOKTANLHYQDLOREVGLK-EINEMLSVLOKEFLHLSKEFATSKDLSAVSQDYFSL 142
 DB 667 LEKRLQERFTDLEKEKSNMEIDMTYQKVIQSLQEEAEHAKATKARLADKNKIYESIE 726
 QY 143 QGFRDNYKGFSLDDEVKNSTEEMRKLFF-----SQEIIADLKGSVASLR 186
 DB 727 EAKSEAMKEMKKLEERTLKQKVENLLEAEKRCSLDCKLQSQOKINELLKOKDVLN 786
 QY 187 BEIRFLT-PLAEVRRRAHNOESLTAEEELKTIRDS---LRDEIGOLSLSKTLTQIA 242
 DB 787 EDVRNLTLKIEQTKRCLTNDLKMOTQOYNTLKMSEKQLQENHLMEMKMLKONA 846
 QY 243 LQKESDLCQIRE 257
 DB 847 ELRKRQDADQMK 861
 RESULT 8
 US-09-067-351-1
 ; Sequence 1, Application US/09067351
 ; Patent No. 5994081
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Baughn, Mariah
 ; TITLE OF INVENTION: HUMAN KERATINS

APPLICANT: Redmond, Mark J.
 APPLICANT: Ijaz, Mohammed K.
 APPLICANT: Parker, Michael D.
 TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
 TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE
 NUMBER OF SEQUENCES: 30
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Morrison & Foerster
 STREET: 545 Middlefield Road, Suite 200
 CITY: Menlo Park
 STATE: CA
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/603,133B
 FILING DATE: 19901025
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Robins, Roberta L.
 REGISTRATION NUMBER: 33,208
 REFERENCE/DOCKET NUMBER: 9313-0004.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 327-7250
 TELEFAX: (415) 327-2951
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 775 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-07-603-133B-15

Query Match 8.5%; Score 113.5; DB 1; Length 775;
 Best Local Similarity 23.2%; Pred. No. 0.042;
 Matches 55; Conservative 45; Mismatches 100; Indels 37; Gaps 9;
 QY 43 LIGTIGLALLGVFLIAPQITIVLLALFITSAGNALYLOKTAN-LHLYQDLOREV 101
 DB 446 LVGLPAPNPSGHEY-YETAGRFSEFILL-----VPSNDDYOTPTMNSVTVRQDLERQLG 498
 QY 102 SLKEINFMLSVLOKEFLHLSKEFATSK-DLSAVSQDYFSCLOGFRDNYKGFESLLDEVK 160
 DB 499 DLRE-----EFNSLSQEIAMTQLDLDLALLFLDMFSM-----FSGIKSTIDAAK 541
 QY 161 NSTEEMRKLFSQE---IIADLKGSVASLRREIRFLTPLAEVRRRAHNOESLTAABEEL 216
 DB 542 SMATKVMKPKRSGLAGTISELTGSLNNAASSISRSISRSISISISISISISISISISIS 597
 QY 217 KTIROSLRDEIGLSQLSKTLTSQIALQKES---SDLCQIRETLSPRKSASPT 270
 DB 598 AGSSDSVNSISQMSAISRLRLREITTTQTEGMNFDISAAVLTKTKIDRSTHISPT 654
 RESULT 7
 US-08-685-576-4
 ; Sequence 4, Application US/08685576
 ; Patent No. 5906819
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaibuchi, Kozi
 ; APPLICANT: Iwamatsu, Akihiro
 ; APPLICANT: Nakano, Takeshi
 ; APPLICANT: Ito, Masaaki
 ; APPLICANT: Takahashi, No. 5906819uaki
 ; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 ; NUMBER OF SEQUENCES: 16

```

;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,351
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANTUT02
; CLONE: 1467090
;
; US-09-067-351-1

```

```

Query Match      8.1%; Score 107.5; DB 2; Length 546;
Best Local Similarity 24.1%; Pred. No. 0.091;
Matches 48; Conservative 31; Mismatches 73; Indels 47; Gaps 7;

QY 80 NALYLQKTANLHLYQDLQREVSLK-EINFMLSVLQKEFLHLSKEFATTSKDL SAVSQDF 138
Db 279 DAAYMSKV-----ELEAKVDALNDEINFRLTNELTELQSQISDTSVVL----- 325
QY 139 YSCLOGFRDNYKG--FESLLDEYKKNSTEMRKLFSEIITA-----DL 178
Db 326 -----MNSRSLLDGIIEVKAQEYEMAKCSRAEAAYQTKFETLQAQAGKHGDDL 378
QY 179 KGSVASLREIEFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTIT 238
379 RNRTRNEISEMNRAIQRLQAEIDNKNQRAKLEAAIAEAE-----RGELALKDARAKQEE 433
QY 239 SQIALQKRESSDLCQIRE 257
Db 434 LEAALQRAK-QDMARQLRE 451

```

```

RESULT 9
US-09-360-490-1
; Sequence 1, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA

```

```

;
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/360,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067,351
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANTUT02
; CLONE: 1467090
;
; US-09-360-490-1

```

```

Query Match      8.1%; Score 107.5; DB 4; Length 546;
Best Local Similarity 24.1%; Pred. No. 0.091;
Matches 48; Conservative 31; Mismatches 73; Indels 47; Gaps 7;

QY 80 NALYLQKTANLHLYQDLQREVSLK-EINFMLSVLQKEFLHLSKEFATTSKDL SAVSQDF 138
Db 279 DAAYMSKV-----ELEAKVDALNDEINFRLTNELTELQSQISDTSVVL----- 325
QY 139 YSCLOGFRDNYKG--FESLLDEYKKNSTEMRKLFSEIITA-----DL 178
Db 326 -----MNSRSLLDGIIEVKAQEYEMAKCSRAEAAYQTKFETLQAQAGKHGDDL 378
QY 179 KGSVASLREIEFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTIT 238
Db 379 RNRTRNEISEMNRAIQRLQAEIDNKNQRAKLEAAIAEAE-----RGELALKDARAKQEE 433
QY 239 SQIALQKRESSDLCQIRE 257
Db 434 LEAALQRAK-QDMARQLRE 451

```

```

RESULT 10
US-08-402-217A-3
; Sequence 3, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,217A
FILING DATE: 10-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PP-0028US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
LIBRARY: mouse
CLONE: GI 53979
US-08-402-217A-3

Query Match      8.0%; Score 106; DB 1; Length 477;
Best Local Similarity 23.9%; Pred. No. 0.1;
Matches 43; Conservative 36; Mismatches 77; Indels 24; Gaps 7;

QY 87 TANLHLYQDLQREVG-SLKEINFMLSVLQKEFLHLSKEFATTSKDLISAYSDQFYSCLQGF 145
Db 160 TAAQLESEQKYNDTAQSLRDVTAQLESEQKYNDTAQSLRDVTAQLESVQEKYNDTAQSL 219
QY 146 RNYKGFESLLDEYKNST----EEMR--KLFQSEIITADLKGVSALREIRFLTPLAEV 199
Db 220 RD-----VSAQLESYKSTLKEIEDLKLNTLQEKVAMAEKSVEDVQOQILTAESTNQY 275
QY 200 RLRAHN-QESLTAIEELTIRSLRDEIGQL-SQLSKTLTSLQIALQKRESSDLCQSIRE 257
Db 276 ARMVQDLQNRSTLKEEIKETSSFLEKITDLKNQL-----ROODEDFRQLEE 324

RESULT 11
US-08-700-178-3
Sequence 3, Application US/08700178
Patent No. 5783669 5700912
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Sellhauer, Jeffrey J.
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/700,178
FILING DATE: August 20, 1996

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,217
FILING DATE: March 10, 1995

```


Search completed: August 13, 2002, 09:24:21
Job time: 152 sec

RESULT 15
US-08-468-577B-12
: Sequence 12, Application US/08468577B
: Patent No. 6001804
: GENERAL INFORMATION:
: APPLICANT: Rabin, Daniel
: TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
: : TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
: : NUMBER OF SEQUENCES: 19
: : CORRESPONDENCE ADDRESS:
: : ADDRESSEE: Spring Kramer Schaefer & Briscoe
: : STREET: 660 White Plains Road
: : CITY: Tarrytown
: : STATE: New York
: : COUNTRY: USA
: : ZIP: 10591-5144
: : COMPUTER READABLE FORM:
: : MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: : COMPUTER: Apple Macintosh
: : OPERATING SYSTEM: System 7.5
: : SOFTWARE: WordPerfect
: : CURRENT APPLICATION DATA:
: : APPLICATION NUMBER: US/08/468,577B
: : FILING DATE: 06-JUN-1995
: : CLASSIFICATION: 514
: : PRIOR APPLICATION DATA:
: : APPLICATION NUMBER: US 08/239,276
: : FILING DATE: 05-MAY-1994
: : PRIOR APPLICATION DATA:
: : APPLICATION NUMBER: US 07/872,646
: : FILING DATE: 08-JUN-1992
: : PRIOR APPLICATION DATA:
: : APPLICATION NUMBER: US 07/715,181
: : FILING DATE: 14-JUN-1991
: : PRIOR APPLICATION DATA:
: : APPLICATION NUMBER: US 07/441,703
: : FILING DATE: 04-DEC-1989
: : PRIOR APPLICATION DATA:
: : APPLICATION NUMBER: US 07/312,543
: : FILING DATE: 17-FEB-1989
: : ATTORNEY/AGENT INFORMATION:

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: H81707
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <DET>
A:Cross-references: GB:AE002306; GB:AE002160; NID:g7190428; PIDN:AAF39253.1; PID:g719043
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0396

Query Match 53.6%; Score 714; DB 2; Length 276;
Best Local Similarity 52.7%; Pred. No. 8.4e-39;
Matches 145; Conservative 58; Mismatches 64; Indels 8; Gaps 2;

1 MTPPTLVIPPS--PPAPSY-----ANRVPOPSLMDKIKKIAATASLILIGTIGFLAL 52
1 MTSPTLVEMPLSCYPPTHTSTACTKRSSSLYPKSLIEIVQVRAAFVSLAILIIGFLAI 60

QY 53 LGLVGLFIAPQITIVLLAFITSLAGNALYLQKTANLHLYQDLQREVSGSLKEINMLSV 112
Db 61 LCHAIGFIAPQVALVLIIVFIISLGNALYLCKTAPRLYLKELQOEVASLKEVNFLLKS 120

QY 113 LQKEFLHLSKEFATSKDLSAVSQDFYSCLOGFRDNKYGFESLIDDEYKNSTEEMRKLFSQ 172
Db 121 VOKEFLGSKDFATSKDLSVSLDFNLLQDFQSSHQGFEDLLEDYKNSAEIDLRLQFSQ 180

QY 173 ELIADLKGVSALREERFRTPLAEVRRRLAHNOESITAAIEELKTIRDSLRDEIGOLSG 232
Db 181 ETVOGLSKTIISLKEEKEIYPLTEEVRRRLAENKEDLLKIQVDLQDIRKLRAEINLSQ 240

QY 233 LSKTITSQIALQKRESSDLCSQIRETSSPKRSAS 267
Db 241 ASKTLSEQIASOIEENKLYANITKALSQSYSDSS 275

RESULT 3
S61491
Inclusion membrane protein A - Chlamydia pneumoniae psittaci
C:Species: Chlamydia pneumoniae psittaci, Chlamydia psittaci
C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 31-Mar-2000
C:Accession: S61491
R:Rockey, D.D.; Heinzen, R.A.; Hackstadt, T.
Mol. Microbiol. 15, 617-626, 1995
A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a protein
A:Reference number: S61491; MUID:95302975
A:Accession: S61491
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <ROC>
A:Cross-references: EMBL:L35036; NID:g516598; PIDN:AAC41443.1; PID:g516599

Query Match 12.2%; Score 162.5; DB 2; Length 355;
Best Local Similarity 21.4%; Pred. No. 0.0025;
Matches 63; Conservative 61; Mismatches 111; Indels 59; Gaps 8;

QY 8 VTPPPAPSYANRV-----PQPSLMDKIKKIAIASLILIGTIGFLALIGHLVG 58
Db 32 IIPISIEAPTSSAAAGAKTAIEPEGRSPLLQICVYLVKIIAIALFVVGIAALVCLVIG 91

QY 59 FLIAPQITIVLLAFITSLA-----GNALYLQKTANLHLYQDLQ-----EV 100
Db 92 SVLSTPSLILMLAILVSVFIVITAIKDTGPSQVVR---HMKQIQOQGEENTRLHTAV 147

QY 101 GSKLEINMLSV--VLOKEFLH-----LSKEFATSKDLSAVSQDFYSCLOGFRDNKYKGE 153
Db 148 ENLKAVNVELSEIQNLKQHLRLSDFGDRLEANTGDTALADFOLSLSEEFKSVGTKYE 207

QY 154 SLIDDEYKNSTEEMRKLFSQIEIADLKGVSALREERFRTPLAEVRRRLAHNOESITAAI 213

Db 208 TMLSPFEKLAGSLAKETFSQEAQVAMSSVTELR-----NLNALKELI 250
QY 214 BELKTRDRLDEIGQLSOLSKTILTSQIALQKRESSDLCSQIRETSSPKRSAS 267
Db 251 TENKTVIEQLKAD----AOLFEEOVRLEKRRKQEBEACSTLSHSIATLOESYT 300

RESULT 4
A72108
Hypothetical protein CP0581 [imported] - Chlamydia pneumoniae (strains CWL029 and
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: A72108
A:Molecule type: DNA
A:Residues: 1-390 <ARN>
A:Cross-references: GB:AE001605; GB:AE001363; NID:g4376455; PIDN:AAD18339.1; PID:g437
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: B81560
A:Molecule type: DNA
A:Residues: 1-390 <REA>
A:Cross-references: GB:AE002217; GB:AE002161; NID:g7189493; PIDN:AAF38399.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CPn0186; CP0581

Query Match 11.6%; Score 154.5; DB 2; Length 390;
Best Local Similarity 20.8%; Pred. No. 0.0092;
Matches 75; Conservative 59; Mismatches 123; Indels 103; Gaps 12;

QY 3 TPTLVIP--PSPAPSYANRVPOPSLMDKIKKIAIASLILIGTIGFLALIGHLVG-- 59
Db 8 TPSAPNIPAPPTPGITPTK--PRSSF--IEKVIIVAKYILFAIATSGALGTIILGLSG 63

QY 60 LIAPQITIVLLAFITSLAGNALYLQKTANLHLYQDLQREVSGSLKEINFMVLYOKEFLH 119
Db 64 ALTPGIGIALLVIFVSWVLGLILKDISGGEERLREEVSRFTSENQRLTV----- 116

QY 120 LSKEFATSKDLSAVSQDFYSCLOGFRD---NYK----- 150
Db 117 ITTLETEVKDLAAKADOLTEAFRNGNGLKTTAEDLEEQVSKLSEOLEALERINQL 176

QY 151 -----GFES-LIDDEYKNSTEEMRKLFSQIEIADLKGVSALREEI 189
Db 177 IQANAGDAQETSELKLIISGWSKVVQEQNTSIQALKVLLQGEWQEQEATHVKAMQEQI 236

QY 190 RFL-----TPLAEVRRRLAHNOESITAAIEELKTIR-----SLRDEIGQLS 231
Db 237 QALQAEILGMHNSOTALQKSVENLLVQDQALTRVVGELLESNKLSOACSLAROEIEKLA 296

QY 232 QLSKTLTSQI-----ALQR-----KESDLCQIRETSSPKRSASPSPT 270
Db 297 QHETSLQORIDAMLAQEQNLAEQVTALEKMKQAKQAESEFIACVDRRTFGRRTPTPTT 356

RESULT 5
B86514
Hypothetical protein CPJ0186 [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: B86514
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;

203 AHNQESLTAABELT*YRDS---LRDEIGQLSQTLSQTALQKRESSDL 251

85 QKTNLHLHYQD-LQREVGSLKEINFWLSVLQKEFLHLSKEFATTSKDLSAVSQDPYSC

Db 668 EKVENLRIMKESLEGEVNSLR---VKLKALENQSPFLIRMSDVEKEISLSIKDLEKLI- 723

QY 144 GFRDNTYKGFESLLDEYKNTEMRKL-----FSQEIADLKGSVASL----- 185

Db 724 -----KEBESLRSEIETDSERKIAETIDETISKKDKDEVAKLKGRIERLKRDRDKLKKALEN 777

QY 186 -----REIRFLTLPLAEVRRLAHNOESLTAAI--EELKTIKIRDSLRDET--GOLSQ 233

Db 778 PEAREVTEKIREVEREIAKREELSVEGKLESLNRLDELIPRKASLEEEIEGLVNKI 837

QY 234 S-----KTLTSQLALQRKSSDLCQIRE 257

Db 838 NALKANINENEALKSLTEKLEKLEKKEGEIYSRIEE 874

RESULT 9

S53348

Alternate names: protein YCL029c

Species: Saccharomyces cerevisiae

R;Variety: strain carlsbergensis

C;Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 06-Feb-1998

C;Accession: S53348; S70387

R;Porter, G.L.

Submitted to the EMBL Data Library, August 1994

A;Description: The polyfunctional HIS4 gene of Saccharomyces carlsbergensis: sequence an

A;Reference number: S53348

A;Accession: S53348

A;Molecule type: DNA

A;Residues: 1-347 <POR>

A;Cross-references: EMBL:U13062; NID:g532779; PID:g532780

R;Porter, G.; Westmoreland, J.; Priebe, S.; Resnick, M.A.

Genetics 143, 755-767, 1996

A;Title: Homologous and homeologous intermolecular gene conversion are not differentia

nd MSH2.

A;Reference number: S70387; MUID:96363910

A;Accession: S70387

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-347 <POW>

A;Cross-references: EMBL:U13062; NID:g532779; PID:g532780

A;Note: the source is designated as Saccharomyces carlsbergensis

C;Genetics:

A;Gene: SGD:B1K1

A;Cross-references: SGD:S0000534; MIPS:YCL029c

A;Map position: 3L

Query Match 9.3%; Score 124; DB 2; Length 347;

Best Local Similarity 27.1%; Pred. No. 0.72;

Matches 56; Conservative 25; Mismatches 86; Indels 40; Gaps 8;

QY 86 KTAHLHYQDLQREVG-SLKEINFMLSVLQKEFLHLSKEFATTSKDLSSVSDQFYSCLQ 144

Db 86 KASNSHVNDHNSVETIELRAQLTIKLRKQLHYKRLDDQRMVLEEQPTF----- 139

QY 145 FRDNTYKGF-----ESLLDEYKNSTE-EMRKLFQSLEIAD-----LKGVSASLUREIRFLTP 194

Db 140 --DYEATIQDREREIDELKQOLEHHRQAKOFFDAENEQLQAVVSQLEHEIK----- 193

QY 195 LAEEVRRLAHQ-----ESLTAIEELKTIKIRDSLRDEIGOLSLSKTLTSQLALQK 246

Db 194 -ENEERNIAHNOPTDSSIEDIELKQVLEOLRSIQDF-----ELHKTWKAREQLK 244

QY 247 ESSDLCQIRETLSPRKSASPSTKSS 273

Db 245 MHNDSLSKEYQNLKSLFTSKPQDSSS 271

RESULT 10

A56539

giantin - human

N;Alternate names: macrogolglin

C;Species: Homo sapiens (man)

C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

C;Accession: A56539; S37536

R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protei

A;Reference number: A56539; MUID:94187728

A;Accession: A56539

A;Molecule type: mRNA

A;Residues: 1-3259 <SEE>

A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715

C;Genetics:

A;Gene: GDB:GOLGB1; GCP; GCP371

A;Cross-references: GDB:454958

A;Map position: 3q13.31-3q13.31

C;Superfamily: giantin

C;Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 9.2%; Score 122.5; DB 1; Length 3259;

Best Local Similarity 23.6%; Pred. No. 13;

Matches 53; Conservative 35; Mismatches 90; Indels 47; Gaps 7;

QY 94 QDLQREVGSLKEINFMLSV-----LQKEFLHLSKEFATTSKDLSSV-----S 135

Db 1584 EKLKVEISLKSKEIAESTEWQEKHELOKEYEILLQSYENVSNSEARIQHVVEAVRQEK 1643

QY 136 ODFYSCLOGFRDNYKGFESLLDEYKNSTEEMRKLF-----SQEIIADLKGSVASLUREI 189

Db 1644 OELYGKLRSTEANKKTEKQLOEAQEQMEEMKMKFAKSKQKILELEENDRLRAEV 1703

QY 190 R-----FLPLAEVRRLAHQESLTAIEELKTIKIRDSLRDEIGOLS-OL 233

Db 1704 HPAGDTAKCMTLLSSNASMKELERVKMEYETLSKTKFQSLMSEKDSUSEEVQDLKHQI 1763

QY 234 SKTLTSQLALQRKSSDLCQIRE--TLSSP-----RKSASPSTK 271

Db 1764 EDNVSKQANLEATEKHNDQNTVTEGTSIPGETEQDSLSMSTR 1808

RESULT 11

VPXRW8

N;Alternate names: hemagglutinin

N;Contains: outer capsid protein VP5; outer capsid protein VP8

C;Species: human rotavirus A

C;Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 01-Mar-1996

C;Accession: F28839

R;Gorziglia, M.; Green, K.; Nishikawa, K.; Taniguchi, K.; Jones, R.; Kapikian, A.2.;

J. Virol. 62, 2978-2984, 1988

A;Title: Sequence of the fourth gene of human rotaviruses recovered from asymptomatic

A;Reference number: A28839; MUID:88275070

A;Accession: F28839

A;Molecule type: genomic RNA

A;Residues: 1-775 <GOR>

C;Genetics:

A;Map position: segment 4

C;Superfamily: rotavirus outer layer protein VP3

C;Keywords: capsid protein; coat protein; glycoprotein; hemagglutinin

F:1-240/Product: outer capsid protein VP8 #status predicted <VP>

F:241-246/Region: cleavage processing #status predicted

F:247-775/Product: outer capsid protein VP5 #status predicted <VP5>

F:32,56,85,97,111,114,132,192,277,324,583,606/Binding site: carbohydrate (Asn) (coval

Query Match 9.2%; Score 122; DB 1; Length 775;

Best Local Similarity 26.8%; Pred. No. 2.5;

Matches 56; Conservative 33; Mismatches 86; Indels 34; Gaps 9;

QY 73 FITSLAGNALYQKTAN-LHLYQDLQREVGSLKKEINFMLSVLQKEFLHLSKEFATTSK-D 130

```

Db      469  FILLVPSNDDYQTPIMNSVTVRQDLERQLGDLRE-----EFSNLSQEIATMTQLID  518
QY      131  LSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMRKLFPSOBIIADLKGSVASLREIR  190
Db      519  LALLPLDMFSN-----PSGIKSTIDAAKSMATVMYKKFKK-----SGLATSISELTR  565
QY      191  FLTPLAEVRR---LAHNOESL---TAAATELKTIRDSLRDELIGLSQLSKYL-TSOLAL  243
Db      566  SLSNAAASSVSRSSSIKSNLSSTVWTDVEQINGSSDSVRNISTOTSIAISKRLREITT  625
QY      244  QRKSS--DLCQSIRETLSPPRKASPSPT  270
Db      626  QTEGMNFIDISA AVLTKIDKQKSTHSPDT  654

```

RESULT 12

giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
Accession: I52300
S:Onda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein
A:Reference number: I52300; MUID:95100974
A:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:G662389; PIDN:BAA05025.1; PID:9808869
A:Superfamily: giantin

Query Match

Query Match	9.1%	Score 121.5;	DB 2;	Length 3225;	
Best Local Similarity	23.6%;	Pred. No. 15;			
Matches 53;	Conservative 35;	Mismatches 90;	Indels 47;	Gaps 7;	
94 QDLQREYGS LKLEINFLMSV-----	LOKEFLHLSEKFA TTSKDSLV-				
1350 EKLVEI ELSKSKIAE	STEQBKHKLEQEKY EILLQSYENV	SNAEIRIQHYV	EA	VRQEK 1609	
136 QDFVSC LQGFDRNYK	GFESLDDYKKNST	BEKRLF-----	SQEITAD LKGS	VASLR E I 189	
1610 QELYGKUR	STEA NKETEKQLQAEQ	EEEM EEMK RFAKSKQK	ITLE	LEENDRLRAEV 1669	
190 R-----	FLPLAEV	ERLHAHNOES	LTAAEELKTT	TRSDRL EIGOLS-QL 233	
1670 HPAGDTAKE	CMETLLSSN	ASMKREELERV	KMETYELSK	FPQSLMSKDSLS	EVQDLKHQI 1729
234 SKTLT	SOIALOR	KESDILCSQ	TR--TLSSP-----	RKSASPTK 271	
1730 EGNV	SQANLEATE	KHDN	GTNTY	EGTSQIPGTE	EQDLSLMSTR 1774

RESULT 13

RESULT 13
T16507
hypothetical protein F59A6.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Cession: T16507
R:Nhan, M.
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of *C. elegans* cosmid F59A6.
A:Reference number: Z18526
A:Accession: T16507
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1286 <NHA>
A:Cross-references: EMBL:U41994; NID:g1123047; PID:g1123050; PIDN:AAA83454.1; CESP:F59A6
C:Genetics:
A:Gene: CESP:F59A6.5
A:Introns: 35/3; 335/3; 685/3; 973/3; 1097/3

Query Match	9.0%	Score 120.5;	DB 2;	Length 1286;
Best Local Similarity	23.3%;	Prod. No. 5.8;		
Matches 49;	Conservative 36;	Mismatches 80;	Indels 45;	Gaps
Qy	92	LYQDLQREVGSGLKEINFMLSVLQKFEFHLHLSKEFATYTKDLSAVSQDYFSC---	LOGFRDN 148	
Db	133	LOETIQRAQDDKKTETTELSSRRSLHVLKELSAKANDIFMYTKDLHDKNEELTSFRME 192		
Qy	149	Y-----KGFSLLDYKNSTE-----MRKLFSGEIIADLKGSAVSLREE 188		
Db	193	YVTKLSEANREKKALEKLEKYKNDMKENDRKSLNLNKEQVITQNVLSE-----VROLSAH 248		
Qy	189	IRFLTPL---AEEVRLAHNQESLTAIEE---LKTIRDSLDEGTGQLSQLSKLTLSOI 241		
Db	249	FEFLTPYRKNAKIRDELVEYHQSAAKVIEESMNDLIKNETLTKELSDNTLVK-----302		
Qy	242	ALQRKSSDLCQSIRETLSPRKSASPSTK 271		
Db	303	-MKNEEDL-----ROTTTASLGDSEOATK 327		

RESULT 14

A:14
 A:12354
 A:12354
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: A12354
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iwano
 Nakazaki, N.; Shimpso, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
 DNA Res. 8, 203-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteri
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: A12354
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-496 <KUR>
 A:Cross-references: GB:BA000019; PTDN:BAB76092.1; PID:g17133529; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alrA393

Query Match

Query Match	9.0%;	Score 120;	DB 2;	Length 496;
Best Local Similarity	21.9%;	Pred. No. 2;		
Matches	57;	Conservative	54;	Mismatches 87; Indels 62; Gaps 11;
Qy	38	IASILIGTIGTFLLAGHLVG-----FLIAPIQTIVILLAFITSLAGNALYLQKT	87	
Dd	8	IAAILILG--GVIAATVGDRIYGRKARLSLFWLRPKNTAVL----VILITGLVLVSATTL	61	
Qy	88	ANLHL-----YQDLQREVSGLSKINFMLSVLQK-----EFLHLSKEFATTSK	129	
Dd	62	AULFIADDEGLRGKVFELEDIQDLQKRE---OLKVAEEOKTOVEIERKNVNOLETTPT	118	
Qy	130	DLSAV-----SODPYSCLQGFRNYVGFSSLDDEYKNSTEEMRKLFQSBI	175	
Dd	119	DKQVETQRDQAQKKELKAQDLAOTQAQYORTSQRLGGVVTVQYAKAEQLSVYNQR--	176	
Qy	176	ADIKGSVASLIREEIRFLTPLABEV-----RRLAHNQEISLTAARELKTIROSLRDGTQL	230	
Dd	177	KALQGAVEQLKTRERLYAEAKATEORDRELANKRAQQAIQEQDRLEANRQALQOORQKI	236	
Qy	231	SQLSKITLTS----QIALQRKE	247	
Dd	237	SOLDKIIQNRRNLIEA-OREE	255	

RESULT 15

15
G82490
methyl-accepting chemotaxis protein VCA0176 [imported] - *Vibrio cholerae* (strain N169)

C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: G82490
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: G82490
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-644 <HEI>
A;Cross-references: GB:AE004358; GB:AE003853; NID:g9657566; PIDN:AAF96089.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0176
A;Map position: 2

Query Match		9.0%	Score 120;	DB 2;	Length 644;
Best Local Similarity		19.2%	Pred. No. 2.7;		
Matches 44;		Conservative 50;	Mismatches 77;	Indels 58;	Gaps 7;
QY	74	ITSLAGNALYLOKTANLHLYQDLQREVGSLKEINFMLSVLQKEFLHLSKEFATTSK---	D 130		
Db	415	ISEIASNAATAETAN-----QASGNADQGRNVVYKAEATSRLAHDIENTGKVVEQ	466		
QY	131	LSAVSODFYSCIQGFR-----DNYKGESLLDEYKN-----S	162		
Db	467	LASTTQETGSLDAIRGISEQTNLLNALNAATEARAGDQGRGFVAVADEVRNLASRTASS	526		
QY	163	TEEMKLFQSEIIADLKGSVASL-----REEIRFLTPLAEVEVRRLAHNOES	208		
Db	527	TEEIOKMINQ-LONDAKNAVSAWDAGKVTTHOGVAASDEAVQVLMISIDRIHIDSDRNTQ	585		
QY	209	LTAATIELKTIRDSRLDEIGQLSQLSKLTLSQIALQKESDLCQIRE	257		
Db	586	VATATEEQSTVVVHTINQNEEINAINETVTS-----TAEELADASKSLRE	630		

Search completed: August 13, 2002, 09:24:56
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:30:42 ; Search time 17.41 Seconds
(without alignments)
607.147 Million cell updates/sec

Title: US-09-673-763-14

Perfect score: 1333

Sequence: 1 MTPTTLIVPPSPAPSYSA.....QIRETLSSPRKASPSSTKSS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128.5	9.6	1957	1 YD86_SCHPO	Q10411 schizosacch
2	124.5	9.3	3210	1 CENF_HUMAN	P49454 homo sapien
3	122	9.2	775	1 VP4_ROTTH1	P11198 human rotav
4	120	9.0	496	1 Y7J3_ANASP	Q05070 anabaena sp
5	116	8.7	776	1 VP4_ROTTHN	P11199 human rotav
6	115	8.6	1325	1 G160_MOUSE	P55937 mus musculus
7	114	8.6	775	1 VP4_ROTTHG	P23045 porcine rot
8	112.5	8.4	622	1 LAM0_DROME	P08928 drosophila
9	112	8.4	391	1 APA4_RAT	P02651 rattus norv
10	111.5	8.4	775	1 VP4_ROTTHM	P11197 human rotav
11	111.5	8.4	1044	1 YAF3_SCHPO	Q09857 schizosacch
12	111.5	8.4	1790	1 US01_YEAST	P25386 saccharomyc
13	111	8.3	432	1 GRAP_HUMAN	P14136 homo sapien
14	110.5	8.3	2104	1 MY53_SCHPO	O14157 schizosacch
15	110	8.3	774	1 VP4_ROTTH	P11200 human rotav
16	110	8.3	978	1 RA50_AQUAE	O67124 aquifex aeo
17	109.5	8.2	776	1 ALM1_SCHPO	Q08778 human rotav
18	109.5	8.2	1727	1 OV71_ONCVO	Q94035 schizosacch
19	109	8.2	432	1 VP4_ROTTH	P31732 onchocerca
20	109	8.2	776	1 TPR_HUMAN	Q06894 canine rota
21	109	8.2	2349	1 K2C8_HUMAN	P12270 homo sapien
22	108.5	8.1	1453	1 Y373_BOVIN	P05787 homo sapien
23	108.5	8.1	504	1 AINX_MOUSE	Q94035 schizosacch
24	108	8.1	804	1 MEA6_HUMAN	P46660 bus musculus
25	108	8.1	1251	1 CYA8_HUMAN	O15320 homo sapien
26	108	8.1	1005	1 RA50_METJA	P40145 homo sapien
27	107.5	8.1	1005	1 HTR1_HALNI	Q58718 methanococc
28	107	8.0	535	1 VP4_ROTTH	P33741 halobacteri
29	107	8.0	776	1 VP4_ROTTH	Q07416 feline rota
30	107	8.0	776	1 VP4_ROTTH	P08895 human rotav
31	106.5	8.0	468	1 K2C7_HUMAN	P08729 human rotav
32	106.5	8.0	729	1 KAR3_YEAST	P17119 saccharomyc
33	106.5	8.0	1189	1 SCII_CHICK	Q09088 gallus gall

34 106.5 8.0 2663 1 CENF_HUMAN Q02224 homo sapien
35 106 8.0 287 1 APAL_MACFA P15568 macaca fasc
36 106 8.0 499 1 AINX_HUMAN Q16352 homo sapien
37 106 8.0 505 1 AINX_RAT P23365 rattus norv
38 106 8.0 775 1 VP4_ROTTH3 P39033 human rotav
39 106 8.0 776 1 VP4_ROTTH1 P12976 simian 11 r
40 106 8.0 776 1 VP4_ROTTH1 P17464 simian 11 r
41 106 8.0 776 1 HMR_MOUSE Q00347 mus musculus
42 106 8.0 794 1 HMR_MOUSE Q00347 mus musculus
43 106 8.0 882 1 MYSP_CAEEL P10567 caenorhabdi
44 105.5 7.9 775 1 VP4_ROTTH5 P30214 human rotav
45 105.5 7.9 775 1 VP4_ROTTH5 P11196 human rotav

ALIGNMENTS

RESULT 1
YD86_SCHPO
ID YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 222.8 kda protein Clf3.06C in chromosome I.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 270690; CAA94624.1; -
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 9.68; Score 128.5; DB 1; Length 1957;
Best Local Similarity 27.48; Pred. No. 1;
Matches 51; Conservative 32; Mismatches 82; Indels 23; Gaps 5;
QY 89 NLHLYQDREVGSLSKEINFMLSVLQKEFLHLSKEFATTSKDLSSVSDYSCGQGFDRN 148
DB 345 NEKLEKLLRNTIGSLKDSRTSNLSOLEEEMVELKESNRTIHSQLT-----DAESKLSFEEQ 400
QY 149 YGFESLLDEYKKNSTEMRKLFSQETIADLKGVSASLUREIRFLTPLAEEVRLAH---N 205
DB 401 NKSUGSIDYQNNLSKDKMKVKQ-----VSSQLEEAR--SSLAHATGKLAETNSE 449
QY 206 QESLTAAEELKTRISRLTSGIQLSOLSKTLTSQALQKESSDLCSTIRETLSSPRKS 265
DB 450 RFOFKKKIKDFKIEQDLRLACLSSNNELKESKLSALDKKQDLNNLRQELKE-----QKK 504
QY 266 ASPSTKSS 273
DB 505 VSESTOSS 512

RESULT 2
CENF_HUMAN

ID AC P49454; Q13246; Q13171; STANDARD; PRT; 3210 AA.
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH antigen).
 GN CENPF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast Carcinoma;
 RX MEDLINE=95348175; PubMed=7542657;
 RA Liao H., Winkfeld R.J., Mack G., Rattner J.B., Yen T.J.;
 RT "CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis."; J. Cell Biol. 130:507-518(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B., Jones D., Yang-Feng T.H., Lee W.-H.;
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression."; Mol. Cell. Biol. 15:5017-5029(1995).
 RN [3]
 RP SEQUENCE OF 2194-3210 FROM N.A.
 RX MEDLINE=95336446; PubMed=7612011;
 RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
 RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal domain sufficient for nuclear localization."; Biochem. Biophys. Res. Commun. 212:220-228(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95370296; PubMed=7642639;
 RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
 RT "The C terminus of mitotin is essential for its nuclear localization, centromere/kinetochore targeting, and dimerization."; J. Biol. Chem. 270:19545-19550(1995).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals interactions with the kinetochore proteins CENP-F and hBUBR1."; J. Cell Biol. 143:49-63(1998).
 CC -!- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
 CC -!- SUBUNIT: HOMO- OR HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS), REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
 CC -!- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
 CC -!- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U19769; AAA82889.1; .
 CC EMBL; U30872; AAA82935.1; .
 CC EMBL; U25725; AAA86889.1; .
 CC HSP; P02649; ILE4.
 CC MIM; 600236; .
 KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;

KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
 FT DOMAIN 14 197
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 273 769
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 823 1328
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1642 1746
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1862 1987
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2207 2568
 FT COILED COIL (POTENTIAL).
 FT REPEAT 2207 2386
 FT 2 X 177 AA TANDEM REPEATS.
 FT REPEAT 2389 2568
 FT 1.
 FT REPEAT 3015 3032
 FT 2.
 FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 16 16
 FT T -> A (IN REF. 2).
 FT CONFLICT 250 250
 FT L -> Q (IN REF. 2).
 FT CONFLICT 272 272
 FT G -> D (IN REF. 2).
 FT CONFLICT 611 611
 FT MISSING (IN REF. 2).
 FT CONFLICT 1494 1589
 FT MISSING (IN REF. 2).
 FT CONFLICT 1611 1611
 FT V -> A (IN REF. 2).
 FT CONFLICT 1811 1811
 FT V -> L (IN REF. 2).
 FT CONFLICT 2242 2243
 FT ER -> DG (IN REF. 3).
 FT CONFLICT 2335 2335
 FT L -> Q (IN REF. 3).
 FT CONFLICT 2492 2492
 FT D -> N (IN REF. 2).
 FT CONFLICT 2545 2545
 FT ELNVAALHNDQAEACK -> SSMREWPQCIIMTKPKVS (IN REF. 3).
 FT SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;

Query Match 9.3%; Score 124.5; DB 1; Length 3210;
 Best Local Similarity 26.2%; Pred. No. 3.4;
 Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;

QY 77 LAGNALYLOK-----TANLHLYQDLOREVGSLKEINFMLSVLOKEFHLHLSKEFAT--- 126
 DB 946 LLETSLSLEKKESSIIISLNKREIELTQENGLKEINASLNQKMNLIQKSEFANYID 1005
 QY 127 -TSKDLGAVS----QDFYSLQGFQFRNYKGFESLLDEYKNSTEEMKRL-----FSGEIIA 176
 DB 1006 EREKSISELSDQYKQKLLILORECEETGNAYEDLSQYKAAQKNSKLECLNNECTSLICE 1065
 QY 177 DLKGSVASLREI-----RFLTPLAEVRLAHNQSLTAAIEELTIRDSLRDEI---- 227
 DB 1066 NRKNELEOLKEAKEHKEHQLTKLAFAEER---NQNLML-----ELETVQOALRSEMTDNO 1118
 QY 228 -----GQLSLSKTLTSLQIALQKSESSDLCSQ 254
 DB 1119 NNSKSEAGGLKQEIIMTLKEQNKMKQKVNLDLQE 1152

RESULT 3
 VP4_ROT1
 ID VP4_ROT1 STANDARD; PRT; 775 AA.
 AC P11198;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
 DE [Contains: Outer capsid proteins VP5 and VP8].
 GN S4.
 OS Human rotavirus (serotype 1 / strain 1076).
 OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
 OX NCBI_TaxID=10944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88275070; PubMed=2839714;
 RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R., Kapikian A.Z., Chanock R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from asymptomatic or symptomatic infections."; J. Virol. 62:2978-2984(1988).
 RN [2]
 RP SEQUENCE OF 1-280.
 RX MEDLINE=86313706; PubMed=3018754;
 RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R., Flores J., Kapikian A.Z., Chanock R.M.;
 RT "Conservation of amino acid sequence of VP8 and cleavage region of

RT 84-Kda outer capsid protein among rotaviruses recovered from
 RL asymptomatic neonatal infection.;
 CC Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
 CC -1- SUBCELLULAR LOCATION: Outer capsid.
 CC -1- PTH: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
 CC OTHER PRODUCT IS VP5.
 CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
 DR PIR: F28839; VPXRW8.
 DR PIR: F28839; VPXRW8.
 DR InterPro: IPR00416; Cap_VP4.
 DR Pfam: PF00426; VP4; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 775
 FT CHAIN 1 240
 FT CHAIN 247 775
 FT CARBOHYD 32 56
 FT CARBOHYD 56 32
 FT CARBOHYD 97 97
 FT CARBOHYD 111 111
 FT CARBOHYD 114 114
 FT CARBOHYD 132 132
 FT CARBOHYD 192 192
 FT CARBOHYD 277 277
 FT CARBOHYD 324 324
 FT CARBOHYD 583 583
 FT CARBOHYD 606 606
 SQ SEQUENCE 775 AA; 87697 MW; EE29A36D36C8018F CRC64;

Query Match 9.2%; Score 122; DB 1; Length 775;
 Best Local Similarity 26.8%; Pred. No. 0.92; Mismatches 33; Conservative 33; Indels 86; Gaps 9;
 Matches 56; Conservative 33; Mismatches 86; Indels 34; Gaps 9;
 QY 73 FITSLAGNALYLQKTN-LHLQDLOREVGLSKNEFMVLSLQKLEFLHLSKEFATSK-D 130
 DB 469 FILLVPSNDYDTPIMNSVTYQDLERQLGLDRE-----EFSLSQEIATQLID 518
 QY 131 LSAVSQDFVSCLOQFRDNYKGFESLLDEYKNSTEMRKLFQSEIADLKGVSASREIR 190
 DB 519 LALLPLDMFMS-----FSGIKSTIDAKSWATVMKFKK-----SGLSATSEISLTR 565
 QY 191 FLTPLAEEVRR---LAHQESL---TAAIEELTIROSLRDEIGQLSLSKTL-TSIAL 243
 DB 566 SLISNAASSVRSKSISSIVNWDVSEQITGSSDSVRNISTQTSASKRLRLRIT 625
 QY 244 QRKES--DLCSQIRETLSSPKSASPST 270
 DB 626 QTEGMNFIDISAVALKTKIDKSTHISPT 654

RESULT 4
 ID V7J3_ANASP STANDARD; PRT; 496 AA.
 AC Q05070;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein alr4393 precursor.
 GN ALR4393.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OC NCBI_TaxID=103690;
 RN [1]
 RP MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yanada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).

RN [2]
 RP SEQUENCE OF 1-166 FROM N.A.
 RX MEDLINE=93308081; PubMed=8391534;
 RA Wei T.-F., Ramasubramanian T.S., Pu F., Golden J.W.;
 RT "Anabaena sp. strain PCC 7120 difa gene encoding a sequence-specific
 RT DNA-binding protein cloned by in vivo transcriptional interference
 RT selection.";
 RL J. Bacteriol. 175:4025-4035(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP003596; BAB76092.1; -;
 DR EMBL; L10036; AAD04182.1; -;
 KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
 FT CHAIN 1 19 POTENTIAL.
 FT SIGNAL 1 19 POTENTIAL.
 FT TRANSMEM 20 496 HYPOTHETICAL PROTEIN ALR4393.
 FT TRANSMEM 45 67 POTENTIAL.
 SQ SEQUENCE 496 AA; 56375 MW; 454D67CF420AFB9E CRC64;
 Query Match 9.0%; Score 120; DB 1; Length 496;
 Best Local Similarity 21.9%; Pred. No. 0.73; Mismatches 57; Conservative 54; Mismatches 87; Indels 62; Gaps 11;
 Matches 57; Conservative 54; Mismatches 87; Indels 62; Gaps 11;
 QY 38 IASLILGTGTFALLGLHVG-----FLIAPQITVLLALFTSLAGNALYLQKT 87
 DB 8 IAILILG--GVATVGDRIQTRVGKARLSLFLNPKNTAVL-----VTILTGLSVSATL 61
 QY 88 ANLHL-----YQDLQREVGLSKNEFMVLSLQK-----EFLHLSKEFATSK 129
 DB 62 AILFIADDEGLRGVFELEDIQDLQKRE---OLKVAEEQKQVETERNKVNQLETTRT 118
 QY 130 DLSAV-----SQDFVSCLOQFRDNYKGFESLLDEYKNSTEMRKLFQSEI 175
 DB 119 DKQVETQRDQAKKEKLKQDLAQOAVQVQTSRLGQVVOYQKAIELASVNR-- 176
 QY 176 ADLKGVSASIREIRFLTPLAEEV-----RLAHQESLTAAIEELKTIROSLRDEIGQL 230
 DB 177 KALQAVEQLKTERRLYAEAKKAEQRDLRANRQAEQRDLRANRQAEQRDLRANRQAEQRDLR 236
 QY 231 SLSKLTLS---QIALQKRE 247
 DB 237 SOLDKTIQNRNLEIA-QREE 255

RESULT 5
 ID VP4_ROTNN STANDARD; PRT; 776 AA.
 AC P11199;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer capsid protein vp4 (Hemagglutinin) (Outer layer protein VP4)
 DE [Contains: Outer capsid proteins VP5 and VP8].
 GN S4.
 OS Human rotavirus (serotype 3 / strain MCN13).
 OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
 OC NCBI_TaxID=10955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88275070; PubMed=2839714;
 RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
 RA Kapikian A.Z., Chanock R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from
 RT asymptomatic or symptomatic infections.";
 RL J. Virol. 62:2978-2984(1988).
 RN [2]

Tue Aug 13 14:23:41 2002

SEQUENCE OF 1-281.
MEDLINE-86313706; PubMed-3018754;
RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
RA Flores J., Kapikian A.Z., Chanock R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection."
RL Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
CC -1- SUBCELLULAR LOCATION: Outer capsid.
CC -1- PPM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC PIR: G286359; VPRW9.
DR PIR: G25904; VPRW9.
DR InterPro: IPR00416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 241
FT CHAIN 248 776
FT CARBOHYD 32
FT CARBOHYD 56
FT CARBOHYD 97
FT CARBOHYD 111
FT CARBOHYD 114
FT CARBOHYD 132
FT CARBOHYD 133
FT CARBOHYD 193
FT CARBOHYD 278
FT CARBOHYD 325
FT CARBOHYD 584
FT CARBOHYD 607
FT CARBOHYD 607
FT CONFLICT 190 190
FT SEQUENCE 776 AA; 87794 MW; CFCADF4DEE986512 CRC64;
Query Match 8.7%; Score 116; DB 1; Length 776;
Best Local Similarity 25.8%; Pred. No. 2.2;
Matches 48; Conservative 31; Mismatches 75; Indels 32; Gaps 7;
QY 94 QDLQREGLSLKEINFMVLSVLOKREFLHLSKEFATSK-DLSAVSODPSYSCLOFRDNKYGK 152
Db 492 QDLERQLGDLRE-----EFNSLSQEIAMTQILDALPLDMFMS-----FSGI 534
QY 153 ESSLDEYKNSTEEMRKLFQSEIITADLKGVSASLREIRFLTLPLAEVRR---LAHQESL 209
Db 535 KSTDAAKSKATKVMKKFKR-----SGLATSISELTRSLNAASSVSRSSIRSNLSI 588
QY 210 --TAAATEELKTIRDSLRDEIGLSQLSKTLTQIALQKRESS--DLCQIREFTLSPRK 264
Db 589 SWTIDVSQIITGSSDSVRNISTQTSASRLRLREITTTQTEGMNFIDISAAVLTKIDKST 648
QY 265 SASPST 270
Db 649 HISPDT 654
RESULT 6
G160_MOUSE STANDARD; PRT; 1325 AA.
ID G160_MOUSE
AC P55937;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
GN GOLGA3 OR MEA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=CD-1; TISSUE=Testis;
RX MEDLINE-97217683; PubMed-9063644;
Kondo M., Sutou S.;
RA "Cloning and molecular characterization of cDNA encoding a mouse
RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
RT autoantigen."
RL DNA Seq. 7:71-82(1997).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
CC DETECTABLE MALE ANTIGEN (SDM).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
CC LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
CC -1- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MEA-1, MEA-19 OR MEA-30 IS THE
CC INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D78270; BAA19612.1;
CC HSSP: P18852; ISCG.
DR MGD; MGI:96958; Golga3.
DR Spermatogenesis; Developmental protein.
KW DOMAIN 201 204 POLY-ALA.
FT DOMAIN 1325 AA; 149880 MW; 3230636962C687B0 CRC64;
FT SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;
Query Match 8.6%; Score 115; DB 1; Length 1325;
Best Local Similarity 21.4%; Pred. No. 4.9;
Matches 48; Conservative 51; Mismatches 83; Indels 42; Gaps 8;
QY 68 VLLALFITSAGNALYLOKTLANHLXQDLQREVSLSKEINFMVLSVLOKREFLHLSKEFAT 127
Db 610 IVLEVALQASDKRELDGRGAR-RLEEDTETSGLLQRLQDLAVKSNQVEHLQOETATL 668
QY 128 SKDLSAVSQDFYS---CLOQFRDNKGFESLDDVYKKNSTE---EMRKLFSQEI----- 175
Db 669 RKQMKVKEQFVQKVMVEAYRRDATSKDQLNELKATKRLDSEMKEL-ROELIKQGE 727
QY 176 -----ADLKGVSASLREIRFLTLPLAEVRRRLAINQESLTA 213
Db 728 KKTVEVHSRLQKMSLVHQMAELEGHLSQVQKE---RDEMEHLSQSLKDFDKQEIATL 784
QY 214 ELKTIIRDSLRDEIGLSQLSKTLTQIALQKRESSDLCQSI 257
Db 785 EANETLKQI-BELOQ--EAKKAITEQKMKRLGSLDLSAQKE 825
RESULT 7
VP4_ROTGP STANDARD; PRT; 775 AA.
ID VP4_ROTGP
AC P23045;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
DE S4.
OS Porcine rotavirus (strain Gottfried).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10917;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90080150; PubMed-2152826;
RA Gorziglia M., Nishikawa K., Hoshino Y., Taniguchi K.;
RT "Similarity of the outer capsid protein VP4 of the Gottfried strain
RT of porcine rotavirus to that of asymptomatic human rotavirus
RL J. Virol. 64:414-418(1990)."

Db 210 TLRVDLENTIQSLRELSFKQIHQSQINESRRRIKQTEYSEIDGRSLSSSEYDAKLKQSLQ 269

QY 184 SL-----REERFLTPLAEEVRRU-----AHNQSLTAABIELTKTIR---DSLKD 225

Db 270 DVRAQYEQMQINDEIQLS--IEDKIORLQEAARTSNSTHKSIEELRSTRVRIDALNA 327

QY 226 EIGOLSQSLKTLTSQIA-LQKSSD 250

Db 328 NINELEQANADLNARIDLERQLDND 353

RESULT 9

AP04_RAT ID AP04_RAT STANDARD; PRT; 391 AA.

AC P02651;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

AP04: Apolipoprotein A-IV precursor (Apo-AIV).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86196059; PubMed=3009456;

RA Boguski M.S., Birkenmeier E.H., Elshourbagy N.A., Taylor J.M.,

RA Gordon J.I.;

RT "Evolution of the apolipoproteins. Structure of the rat apo-A-IV gene

RT and its relationship to the human genes for apo-A-I, C-II, and E";

RL J. Biol. Chem. 261:6398-6407(1986).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=84298074; PubMed=6591177;

RA Boguski M.S., Elshourbagy N.A., Taylor J.M., Gordon J.I.;

RT "rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino

RT acid segment with amphipathic helical potential";

RL Proc. Natl. Acad. Sci. U.S.A. 81:5021-5025(1984).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=87008540; PubMed=3020028;

RA Haddad I.A., Ordoas J.M., Fitzpatrick T., Karathanasis S.K.;

RT "Linkage, evolution, and expression of the rat apolipoprotein A-I, C-

RT III, and A-IV genes";

RL J. Biol. Chem. 261:13268-13277(1986).

CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND

CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN

CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR

CC COMPONENT OF HDL AND CHYLOMICRONS.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: SECRETED IN PLASMA.

CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH

CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-

CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-

CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY

CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL

CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.

CC -1- SIMILARITY: BELONGS TO THE APOA4 / APOA4 / APOE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M00002; AAA85909.1; -

CC DR EMBL; J02588; AAA40747.1; -

CC DR EMBL; M13508; AAA40748.1; -

CC DR PIR; A03095; LPRTA4.

CC DR PIR; A25214; A25214.

PIR: C24700; C24700.

DR HSSP; P02671; IFZB.

DR InterPro; IPR000074; Apolipoprotein.

DR Pfam; PF01442; Apolipoprotein; 1.

KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.

FT SIGNAL 1 20

FT CHAIN 21 391

FT DOMAIN 33 330

FT REPEAT 33 54

FT REPEAT 60 81

FT REPEAT 82 103

FT REPEAT 115 136

FT REPEAT 137 158

FT REPEAT 159 180

FT REPEAT 181 202

FT REPEAT 203 224

FT REPEAT 225 246

FT REPEAT 247 268

FT REPEAT 269 286

FT REPEAT 287 308

FT REPEAT 309 330

FT REPEAT 374 385

FT DOMAIN 253 253

FT VARIANT 391 AA; 44456 MW; 24095004A809201D CRC64;

SQ SEQUENCE 391 AA; 44456 MW; 24095004A809201D CRC64;

Query Match 8.4%; Score 112; DB 1; Length 391;

Best Local Similarity 21.4%; Pred. No. 1.8; 87; Indels 74; Gaps 13;

Matches 59; Conservative 56; Mismatches 56;

QY 59 FLIAPQITVLLAL-----FTSLAGNA-----L 82

Db 2 FLKAVLTVALVATGTAQAEVTSQVAVMDYFTQLSNNKAQVEQLQKTDVTOQLNTL 61

QY 83 YLOKTANLHLY-QDLQREVGLSKEINFLSV---LQKFLHLSKEFATTSKDLSA----- 133

Db 62 FQDKLGNINTYADLQN-----KLVPFAVQLSGHLTKETKRVREIQLKELEDRANMMPH 116

QY 134 ---VSQDFYSCLOQFRNYKGFESILDEYKNS-TEEMRKL---FSQEIADLKGSVASLR 186

Db 117 ANKVSQMGFQVQKQLEHLRPIYATDLQAIQAQTDQMKRQLTPIQRMOTTIQDNVENLQ 176

QY 187 BEIRFLTPLAEVR-RLAHNQE---SLTAAIEELK-TIRSDRLDEICOLSLSKLTLSQ 240

Db 177 SS---MVFPANELKEKFNQNMGLKQLTTPRANELKATIDQNLLEDRSLRLAPLAEGVOEK 233

QY 241 IALQ-----RKESDLCQIRETSSPRKSASP 268

Db 234 LNHOMEGLAFOFMKNAEELQTKVSTNIDQLQKNLAP 269

RESULT 10

VP4_ROTUM STANDARD; PRT; 775 AA.

AC P11197;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)

DE [Contains: Outer capsid proteins VP5 and VP8].

GN S4.

OS Human rotavirus (serotype 1 / strain M37).

OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.

OX NCBI_TaxID=10954;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88275070; PubMed=2839714;

RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,

RA Kpiklian A.Z., Chanock R.M.;

RT "Sequence of the fourth gene of human rotaviruses recovered from

RT asymptomatic or symptomatic infections.";

RL J. Virol. 62:2978-2984(1988).

RN [2]

```

RA  MEDLINE=86313706; PubMed=3018754;
RX  Flores J., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
RA  Flores J., Kapikian A., Chanock R.M.;
RT  "Conservation of amino acid sequence of VP8 and cleavage region of
RT  84-kDa outer capsid protein among rotaviruses recovered from
RT  asymptomatic neonatal infection.";
RL  Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
CC  -1- SUBCELLULAR LOCATION: Outer capsid.
CC  -1- PTM: VP8 IS ONE OF TWO TRIPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC  OTHER PRODUCT IS VP5.
CC  -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
DR  PIR: E28839; VPKRW7.
DR  InterPro: IPR000416; Cap_VP4.
DR  Pfam: PF00426; VP4; 1.
KW  Coat protein; Glycoprotein.
FT  CHAIN 1 775
FT  CHAIN 1 240
FT  CHAIN 1 775
FT  CARBOHYD 32 32
FT  CARBOHYD 56 56
FT  CARBOHYD 97 97
FT  CARBOHYD 111 111
FT  CARBOHYD 114 114
FT  CARBOHYD 132 132
FT  CARBOHYD 146 146
FT  CARBOHYD 192 192
FT  CARBOHYD 277 277
FT  CARBOHYD 324 324
FT  CARBOHYD 583 583
FT  CARBOHYD 606 606
SQ  SEQUENCE 775 AA; 87402 MW; 05CA6DDAA258550B CRC64;

Query Match      8.4%; Score 111.5; DB 1; Length 775;
Best Local Similarity 23.2%; Pred. No. 4.4;
Matches 43; Conservative 38; Mismatches 75; Indels 29; Gaps 6;

QY  94 QDLQREVGSLKEINFMVLSVQKEFLHLSKEFATSK-DLSAVSQDFYSCLOQFRDNYKGF 152
DB  491 QDLERQLGDLRE-----EFNSQIEIAMTQLDIDLALLPLDMFSM-----FSGI 533

QY  153 ESLLDYKYNSTEMRKLFQSE-----IADLKGSVASLREIRFLPLAEVRRLAHNOESLTA 208
DB  534 KSTIDAKSNMAYKMKFRSGLATGISELTGSLNNAASSISRSISNIS-----V 589

QY  209 LFAAIELKTRIDSLDEIGLSQSLSKTLTSQIALQKES---SDLCQIRETLSSPRKS 265
DB  590 WTDVSRQIAGSDSVSNISTQMSAISRRRLRLREITQTQEGMNFDDISRAVLKTKIDRSTH 649

266 ASPST 270
650 ISPDT 654

RESULT 11
ID  YAF3_SCHPO
AC  Q09857; Q997V0; STANDARD; PRT; 1044 AA.
DT  01-FEB-1996 (Rel. 33, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DE  Hypothetical 119.1 kDa protein C29E6.03C in chromosome I.
GN  SPAC29E6.03C OR SPAC30.07C.
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomycetes.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=972;
RA  Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,

```

```

RA  Rajandream M.A., Walsh S.V.;
RL  Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=972;
RA  McDougall R.C., Rajandream M.A., Barrell B.G., Brown S., Murphy L.,
RA  Jones L., McNeil A., Harris D.;
RL  Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC  -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; 266525; CA91425.1; ALT_SEQ.
DR  EMBL; AL136538; CAB66466.1; -.
KW  Hypothetical protein; Coiled coil.
FT  DOMAIN 600 1014 COILED COIL (POTENTIAL)
SQ  SEQUENCE 1044 AA; 119127 MW; 52ACE0E174725A59 CRC64;

Query Match      8.4%; Score 111.5; DB 1; Length 1044;
Best Local Similarity 21.2%; Pred. No. 6.2;
Matches 41; Conservative 39; Mismatches 78; Indels 35; Gaps 4;

QY  94 QDLQREVGSLKEINFMVLSVQKEFLHLSKEFATSKDLSAVSQDFYSCLOQFRDNYKGF 153
DB  729 RDLLESEKSLNLSLALESKNKKLENDLNLTEKLNKNADTESFKNTIREAELSKK 788

QY  154 SLLDEYKYNSTEMRKLFQSEIIADLKGSVASLREIRFLPLAEVRRLAHNOESLTA 213
DB  789 ALNDNLGN-----KENIISDLKNK-----LSESTRRLQELQSLNQDK 826

QY  214 RELKTIROSL---RDEICQLSLSKTLTSQIALQKESDLCQIR-----ETLS 260
DB  827 NOIEFLNIRISAADELSSMESINKNOANELKAKQKCNLQEKINFGNKLAKHEKTEKIS 886

QY  261 SPRKSASPSTKSS 273
DB  897 SLEKDLAATKTA 899

RESULT 12
ID  USO1_YEAST
AC  P25386; STANDARD; PRT; 1790 AA.
DT  01-MAY-1992 (Rel. 22, Created)
DT  01-MAY-1992 (Rel. 22, Last sequence update)
DE  16-Oct-2001 (Rel. 40, Last annotation update)
DE  Intracellular protein transport protein USO1.
GN  USO1 OR INT1 OR YDL058W.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=X2180-1A;
RX  MEDLINE=91185402; PubMed=2010462;
RA  Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA  Yamasaki M.;
RT  "A cytoskeleton-related gene, usol, is required for intracellular
RT  protein transport in Saccharomyces cerevisiae.";
RL  J. Cell Biol. 113:245-260(1991).
RN  [2]
RP  SEQUENCE OF 782-1790 FROM N.A.
RA  Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA  Kendrick K.E.;
RL  Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.

```

[3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X54378; CAA38253.1; -
CC ENBL; L03188; AAB00143.1; -
CC ENBL; U53668; AAB66659.1; -
CC PIR; A38455; A38455.
CC HSP; P80220; IDIP.
CC SGD; S0002216; USOL.
CC InterPro; IPR002017; Spectrin.
CC TransPort; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
CC DOMAIN 1 724 GLOBULAR HEAD.
CC DOMAIN 725 1790 COILED COIL (POTENTIAL).
CC DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
CC DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
CC DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
CC CONFLICT 847 847 G -> E (IN REF. 2).
CC CONFLICT 924 924 E -> K (IN REF. 2).
CC CONFLICT 1253 1253 V -> I (IN REF. 2).
CC CONFLICT 1319 1319 I -> V (IN REF. 2).
CC CONFLICT 1461 1461 N -> S (IN REF. 2).
CC CONFLICT 1581 1581 G -> S (IN REF. 2).
CC CONFLICT 1600 1600 I -> V (IN REF. 2).
CC CONFLICT 1661 1661 R -> S (IN REF. 2).
CC CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
CC SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
Query Match 8.48; Score 111.5; DB 1; Length 1790;
Best Local Similarity 22.38; Pred. No. 12;
Matches 54; Conservative 42; Mismatches 81; Indels 65; Gaps 9;
Oy 74 ITSAGNALYLQKLTANLHLYQDLOREVGSL-----KEINFMVLVLQKFLHLSKE-- 123
Db 964 LKSLANN-----YKMQAENESLIKAVEESKNSSIQLSNQLNKIDMSQKE 1011
Oy 124 -----FATTSKDLNVSODFYSCLOGFDNKGESLL-----D 157
Db 1012 NFQIERGSIENIEOLKLTISDLETKETIKSDSSKDEYESQISLKEKLETATTAND 1071
Oy 158 EYKNSTEEMKRLFSOEITADLKGVSASLREIRFLTPLAEE-VRRLAHNQESL----- 209
Db 1072 ENVNKISLTKT-RELEAEAL-AAYKNLKNLETKLETSEKALKEVKEEHLKEEKTLQ 1129
Oy 210 -----TAAIEELTKIR---DSLREDEIGQLSOLSKTLTQSIALQKRESSDLCQSOIRETLSP 262
Db 1130 EKEATETKOQLNSLRANLESLEKEHDLAAQLKYEEQIANKEQYNEISQINDEITST 1189
Oy 263 RK 264
Db 1190 QQ 1191
RESULT 13

GFAP_HUMAN STANDARD; PRT; 432 AA.
ID GFAP_HUMAN
AC P14136;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glial fibrillary acidic protein, astrocyte (GFAP).
GN GFAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89296968; PubMed=2740350;
RA Reeves S.A., Helman L.J., Allison A., Israel M.A.;
RT "Molecular cloning and primary structure of human glial fibrillary
RT acidic protein."; Proc. Natl. Acad. Sci. U.S.A. 86:5178-5182(1989).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90294716; PubMed=2163003;
RA Brenner M., Lampel K., Nakatani Y., Mill J., Banner C., Mearow K.,
RA Dohadwala M., Lipsky R., Freese E.;
RT "Characterization of human cDNA and genomic clones for glial
RT fibrillary acidic protein."; Brain Res. Mol. Brain Res. 7:277-286(1990).
RL [3]
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92343304; PubMed=1636374;
RA Kumanishi T., Usui H., Ichikawa T., Nishiyama A., Katagiri T.,
RA Abe S., Yoshida Y., Washiyama K., Kuwano R., Sakimura K.;
RT "Human glial fibrillary acidic protein (GFAP): molecular cloning of
RT the complete cDNA sequence and chromosomal localization (chromosome
RT 17) of the GFAP gene."; Acta Neuropathol. 83:569-578(1992).
RL [4]
RN [4]
RP SEQUENCE OF 352-417 FROM N.A.
RX MEDLINE=89386721; PubMed=2780570;
RA Duguid J.R., Bohmont C.W., Liu N.G., Tourtellotte W.W.;
RT "Changes in brain gene expression shared by scrapie and Alzheimer
RT disease."; Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).
RL [5]
CC -!- FUNCTION: GFAP, A CLASS-III INTERMEDIATE FILAMENT, IS A CELL-
CC SPECIFIC MARKER THAT, DURING THE DEVELOPMENT OF THE CENTRAL
CC NERVOUS SYSTEM, DISTINGUISHES ASTROCYTES FROM OTHER GLIAL CELLS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; J04569; AAA52528.1; -
CC ENBL; S40719; AAB22581.1; -
CC ENBL; M26638; AAA52529.1; -
CC PIR; A32936; A32936.
CC PIR; A60052; A60052.
CC HSC-2DPAGE; P14136; HUMAN.
CC MIN; 137780; -
CC InterPro; IPR001664; IF.
CC Pfam; PF00038; Filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil.
CC DOMAIN 1 72 HEAD.
CC DOMAIN 73 377 ROD.
CC DOMAIN 378 432 TAIL.
CC DOMAIN 73 104 COIL 1A.
CC DOMAIN 105 115 LINKER 1.
CC DOMAIN 116 214 COIL 1B.
FT

```

FT DOMAIN 215 230 LINKER 12.
FT DOMAIN 231 252 COIL 2A.
FT DOMAIN 253 256 LINKER 2.
FT DOMAIN 257 377 COIL 2B.
SQ SEQUENCE 432 AA; 49880 MW; E6C3B3454C3F1250 CRC64;

Query Match
Best Local Similarity 8.3%; Score 111; DB 1; Length 432;
Matches 60; Conservative 34; Mismatches 86; Indels 66; Gaps 11;

QY 82 LYLOQTANHLXQ-----DLQREVSL-KEINFMLSVLQKFLHLSKEFA--TT 127
DQ 161 LRLAENNLAAVROEADATLARLDLKRKIESLEEFRLKRIHEEYRELOEQARQOV 220
QY 128 SKDLSAVSDYFSCLOQFRDNYKGFES-----LLDEYKN 161
DQ 221 HVLDVAPKPDLTAAKEIRTOYEMASSNMHEAEWYRSKFADLTDAARNALLRQAKH 280
QY 162 STEEMKLFSGEITAD---LKGVSASLREEIRFLTPLAEEVRLAHNOESLTAIEELKT 218
DQ 281 EANDYRROL-QSLTCDLSLGTNESLERQMR--EQEERHVREAASQEARLEEEGQS 337
QY 219 IRDSLRL---DEIGQLSLSKLTLSQIALQK-----ESSDLCQSIQRETLS 261
DQ 338 LKDEMARHLQEQDQLNLYKALDIEIATYRKLEGEENRITIPVQTFSLN--QIRET-SL 394
QY 262 PKRSAS 267
DQ 395 DTKSVS 400

```

```

RESULT 14
MYS3_SCHPO
ID MYS3_SCHPO STANDARD; PRT; 2104 AA.
AC O14157; 042730;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin type II heavy chain 2.
GN MYO3 OR MYP2 OR SPAC4A8.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98119390; PubMed=9459302;
Moteji F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
"Identification of Myo3, a second type-II myosin heavy chain in the
fission yeast Schizosaccharomycetes pombe.";
FEBS Lett. 420:161-166(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98062346; PubMed=9398685;
Bezanilla M., Forsburg S.L., Pollard T.D.;
"Identification of a second myosin-II in Schizosaccharomycetes pombe:
Myo2p is conditionally required for cytokinesis.";
Mol. Biol. Cell 8:2693-2705(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Skelton J.; Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING
CC THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN
CC CONJUNCTION WITH MYO2.
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB007633; BAA24579.1; -
DR EMBL; AF029788; AAC04615.1; -
DR EMBL; Z98762; CAB11475.1; -
DR HSPSP; P08799; JMN0.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
KW Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation.
FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 182 189 ATP (POTENTIAL).
FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1193 1193 D -> G (IN REF. 2).
FT CONFLICT 1304 1304 E -> G (IN REF. 2).
FT CONFLICT 1344 1344 E -> K (IN REF. 2).
FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A05489594028D258 CRC64;

Query Match
Best Local Similarity 8.3%; Score 110.5; DB 1; Length 2104;
Matches 41; Conservative 48; Mismatches 72; Indels 27; Gaps 6;

QY 93 YODLQREVGLSKAEINFMLSVLQKFLHLSKEFATTSKLSAVSQDYFSCLOQFRDNYKGF 152
DQ 918 YEFLLAEKOSIEE---DLANKQTEISYLSLSLSTLEKLSIKKDEQTISSKYKELEKDY 974
QY 153 ESLDEYKYNSTEEMRKLFQSOEIIADLKGVSASLREEIRFLTPLAEEVRLAHNOESLTA 212
DQ 975 LNMADYQHSSQHLNLSL--EKAINENLNIRELNKRLDLD-----ELLKQKSYDTK 1026
QY 213 IEELATIRDSLDEI---GQL-----SOLSKTLFSQLQKESSDLCQSIQRE 257
DQ 1027 VOELREENASLKQCRITYESQLASLVSKYSETESELNKK-EAELVIFQKEITEYRDQLHK 1085
QY 258 TLSSPRKS 265
DQ 1086 AFQNPKEK 1093

RESULT 15
VP4_ROTHT
ID VP4_ROTHT STANDARD; PRT; 774 AA.
AC P11200;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Human rotavirus (serotype 4 / strain St. Thomas 3).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10960;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88275070; PubMed=2839714;
Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
Kapikian A.Z., Chanock R.M.;
"Sequence of the fourth gene of human rotaviruses recovered from
RT asymptomatic or symptomatic infections.";
RL J. Virol. 62:2978-2984(1988).
RN [2]

```

RP SEQUENCE OF 1-280.
RX MEDLINE=86313706; PubMed=3018754;
RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
RA Flores J., Kapikian A.Z., Chanock R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection";
RL Proc Natl Acad Sci U.S.A. 83:7039-7043(1986).
CC -1- SUBCELLULAR LOCATION: Outer capsid.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; glycoprotein.
FT CHAIN 1 774 OUTER CAPSID PROTEIN VP4.
FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.
FT CHAIN 247 774 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 774 AA; 87455 MW; D397E5C6125A8FA8 CRC64;

Query Match 8.3%; Score 110; DB 1; Length 774;
Best Local Similarity 25.3%; Pred. No. 5.4; Indels 32; Gaps 7;
Matches 47; Conservative 31; Mismatches 76; Indels 32; Gaps 7;
QY 94 QDLOREVGSLKEINFMLSVLQKEFLHLSKEFATTSK-DLSAVSQDFYSCLOGFRDNYKGF 152
Db 491 QDLERQLGLDRE-----EFNSLSQEIATITQIDLALLPLDMFSM-----FSGI 533
QY 153 ESLDDEYKNSTEMRKLFQSQEIADLKGVSASLREIRFLTPLAEEVRR--LAHQESL 209
Db 534 KSTIDAAKSMATKVMKKFR-----SGLATSISETLRSLSNAASSVSRSSIRSNISSI 587
QY 210 ---TAAIEELKTIROSLRDEIGLSQSLKTLTSQIALQKES--DLCSQIRETLSSPRK 264
Db 588 SEWTDVSEQIAGSSDSVRNISTQTSASRRLRLREITTTQEGMNFIDISAAVLKTKIDRST 647
QY 265 SASPST 270
Db 648 HIRPDT 653

Search completed: August 13, 2002, 09:30:44
Job time: 380 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:30:16 ; Search time 48.57 Seconds
(without alignments)
972.362 Million cell updates/sec

Title: US-09-673-763-14
Perfect score: 1333
Sequence: 1 MTTPTLIVPPSPAPSYSA.....QIRETLSPKSPSTKSS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17299429 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1333	100.0	273	2	O69196
2	1328	99.6	273	2	Q9A5A5
3	1318	98.9	273	2	Q9AMB2
4	1315	98.6	273	2	Q9AMA4
5	1311	98.3	273	16	O84121
6	1310	98.3	273	2	Q9AMA6
7	1308	98.1	273	2	Q9AMA9
8	1307	98.0	273	2	Q9AMA7
9	1306	98.0	273	2	Q9AFK9
10	1306	98.0	273	2	Q9A056
11	1305	97.9	273	2	Q9A056
12	1305	97.9	273	2	Q9A056
13	1302	97.7	273	2	Q9A056
14	1293	97.0	273	2	Q9A056
15	1292	96.9	273	2	Q9A056
16	1209.5	90.7	256	2	Q9F7L1

17	962	72.2	197	2	Q9F7T6
18	756.5	56.8	158	2	Q9F7L0
19	714	53.6	276	16	Q9PKR8
20	612	45.9	174	2	Q9F7L2
21	582	43.7	121	2	Q9AM93
22	564	42.3	174	2	Q9F7K8
23	503	37.7	109	2	Q9F7W7
24	162.5	12.2	355	2	Q46210
25	154.5	11.6	390	16	Q92828
26	128.5	9.6	472	13	Q07962
27	126	9.5	1179	17	O59462
28	124	9.3	263	13	Q98TG6
29	124	9.3	347	3	Q02428
30	123.5	9.3	471	13	P87360
31	122.5	9.2	3259	4	Q14789
32	122	9.2	24	2	Q9F7T8
33	122	9.2	775	12	Q86200
34	122	9.2	1931	5	Q9NCF9
35	120.5	9.0	511	13	Q90ZF7
36	120.5	9.0	718	4	Q9BY13
37	120.5	9.0	1286	5	Q21025
38	120	9.0	644	16	Q9KX20
39	120	9.0	1931	5	Q9VXH9
40	119.5	9.0	406	2	Q9L514
41	119.5	9.0	1743	5	O96063
42	119	8.9	897	13	O13098
43	117.5	8.8	775	12	O90314
44	117.5	8.8	1475	5	O76447
45	117	8.8	1101	4	Q96SB9

ALIGNMENTS

RESULT 1

O69196 PRELIMINARY; PRT; 273 AA.
ID O69196;
AC O69196;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LGV-434;
RX MEDLINE=99043938; PubMed=9826388;
RA Bannantine J.P., Stamm W.E., Suchland R.J., Rockey D.D.;
RT "Chlamydia trachomatis Inca is localized to the inclusion membrane and
is recognized by antisera from infected humans and primates.";
RL Infect. Immun. 66:6017-6021(1998).
DR EMBL; AF067958; AAC82641.1;
SQ SEQUENCE 273 AA; 30272 MW; E64B846A5D6824EC CRC64;

Query Match 100.0%; Score 1333; DB 2; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILGTIGTGLLGHVGLF 60
Db	1	MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILGTIGTGLLGHVGLF 60
Qy	61	IAPOTTIVLLALFTTSLAGNALYLOKNTANLHLYODLQREVGSLKEINFMLSVLQKEFLH 120
Db	61	IAPOTTIVLLALFTTSLAGNALYLOKNTANLHLYODLQREVGSLKEINFMLSVLQKEFLH 120
Qy	121	SKEFATTSKDLSSVQDFYSCLOGFRDNYKGFESLLDEYNKSTEEMRKLFQSQIIADLKG 180
Db	121	SKEFATTSKDLSSVQDFYSCLOGFRDNYKGFESLLDEYNKSTEEMRKLFQSQIIADLKG 180

SEQUENCE FROM N.A.

RC STRAIN-TW-5;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF326993; AAG61090.1; -.
SQ SEQUENCE 273 AA; 30318 MW; 22729569405E422B CRC64;

Query Match 98.9%; Score 1318; DB 2; Length 273;
Best Local Similarity 98.9%; Pred. No. 1.6e-80;
Matches 270; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120
QY 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKS 180
QY 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
Db 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
QY 241 IALQKESDLCQIRETLSPRKSPSTKSS 273
Db 241 IALQKESDLCQIRETLSPRKSPSTKSS 273

RESULT 4

Q9AMA4 PRELIMINARY; PRT; 273 AA.
AC Q9AMA4; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-404-L;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327010; AAG61107.1; -.
SQ SEQUENCE 273 AA; 30346 MW; 7E9F9569405E5469 CRC64;

Query Match 98.6%; Score 1315; DB 2; Length 273;

Best Local Similarity 98.5%; Pred. No. 2.5e-80;
Matches 269; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120

QY 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
Db 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
QY 241 IALQKESDLCQIRETLSPRKSPSTKSS 273
Db 241 IALQKESDLCQIRETLSPRKSPSTKSS 273

RESULT 2

Q9AMA5 PRELIMINARY; PRT; 273 AA.
AC Q9AMA5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-434/BU;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327009; AAG61106.1; -.
SQ SEQUENCE 273 AA; 30259 MW; E0B3496A529B14EC CRC64;

Query Match 99.6%; Score 1328; DB 2; Length 273;

Best Local Similarity 99.6%; Pred. No. 3.4e-81;
Matches 272; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALGHVGL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVSGSLKEINFMVLSVLOKEFLHL 120
QY 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
QY 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
Db 181 SVASLREIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSLSKLTTSQ 240
QY 241 IALQKESDLCQIRETLSPRKSPSTKSS 273
Db 241 IALQKESDLCQIRETLSPRKSPSTKSS 273

RESULT 3

Q9AMB2 PRELIMINARY; PRT; 273 AA.
AC Q9AMB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]

```
QY 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
|||||
Db 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKS 180
|||||
QY 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
Db 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
|||||
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
|||||

RESULT 5
O84121 PRELIMINARY; PRT; 273 AA.
AC O84121;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA OR CT119.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RA MEDLINE=99000809; PubMed=9784136;
RX Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Itatou R.L., Zhao Q., Koonin E.V., Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RL Chlamydia trachomatis."
RN Science 282:754-759(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9301;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9301;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=SA-1, APACHE-2, IOL-238, UW-36, 2B, 10A, AND 11A;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Ossewaarde J.M., Morre S.A., van den Brule A.J.C., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia trachomatis isolates with the Inca I47T mutation."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF201286; AAC67710.1;
DR EMBL; AF279346; RAG31466.1;
DR EMBL; AF327329; RAG11233.1;
DR EMBL; AF326992; AAG61089.1;
DR EMBL; AF326994; AAG61091.1;
DR EMBL; AF326998; AAG61095.1;
DR EMBL; AF327002; AAG61099.1;
DR EMBL; AF327004; AAG61101.1;
DR EMBL; AF327006; AAG61103.1;
DR EMBL; AF327012; AAG61109.1;
DR EMBL; AF327014; AAG61111.1;
DR EMBL; AF327015; AAG61112.1;
KW Complete proteome.
SQ SEQUENCE 273 AA; 30327 MW; 6945AE7E8B8BEBF3 CRC64;
```

```
Query Match 98.3%; Score 1311; DB 16; Length 273;
Best Local Similarity 98.2%; Pred. No. 4.6e-80;
Matches 268; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MTTPTLIIVPPSPAPSYSANRVPOPSLMDKIKKTAATASLILIGTIGFALLGHLVGFL 60
|||||
Db 1 MTTPTLIIVPPSPAPSYSANRVPOPSLMDKIKKTAATASLILIGTIGFALLGHLVGFL 60
|||||
QY 61 IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGLSKNEINFMLSVLQKEFLHL 120
|||||
Db 61 IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGLSKNEINFMLSVLQKEFLHL 120
|||||
QY 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
|||||
Db 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
|||||
QY 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
Db 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
|||||
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
|||||

RESULT 6
O9AMA6 PRELIMINARY; PRT; 273 AA.
AC O9AMA6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=440-L;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Ossewaarde J.M., van den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia trachomatis isolates with the Inca I47T mutation."
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL; AF327008; AAG61105.1;
SQ SEQUENCE 273 AA; 30465 MW; 2267957C55E5E72B CRC64;

Query Match 98.3%; Score 1310; DB 2; Length 273;
Best Local Similarity 98.5%; Pred. No. 5.4e-80;
Matches 269; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MTTPTLIIVPPSPAPSYSANRVPOPSLMDKIKKTAATASLILIGTIGFALLGHLVGFL 60
|||||
Db 1 MTTPTLIIVPPSPAPSYSANRVPOPSLMDKIKKTAATASLILIGTIGFALLGHLVGFL 60
|||||
QY 61 IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGLSKNEINFMLSVLQKEFLHL 120
|||||
Db 61 IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGLSKNEINFMLSVLQKEFLHL 120
|||||
QY 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
|||||
Db 121 SKEFATSKDLSAVSODFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKS 180
|||||
QY 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
Db 181 SVASLREIRFLTPLAEEVRRRLAHNQSLSLTAATEELKTTIRDSLRDEIGOLSQSLTTSQ 240
|||||
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
|||||
```

0;

```

Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATAEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTVVIEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
QY 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273
Db 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273

RESULT 10
Q9AMA7
ID Q9AMA7 PRELIMINARY; PRT; 273 AA.
AC Q9AMA7;
QY 01-JUN-2001 (TrEMBLrel. 17, Created)
QY 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
QY 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-31;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
RT trachomatis isolates with the Inca I47T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327007; AAG61104.1; -.
SQ SEQUENCE 273 AA; 30337 MW; 6945AE7E8B8824F3 CRC64;

```

```

Query Match 98.0%; Score 1306; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 9.9e-80;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
QY 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATAEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTVVIEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
QY 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273
Db 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273

```

```

RESULT 11
Q99Q56
ID Q99Q56 PRELIMINARY; PRT; 273 AA.
AC Q99Q56;
QY 01-JUN-2001 (TrEMBLrel. 17, Created)
QY 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

```

DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WT9329, AND MT9309;
RC Viratyosin W., Rocky D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia
RT trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF327331; AAK11235.1; -.
DR EMBL: AF327330; AAK11234.1; -.
SQ SEQUENCE 273 AA; 30339 MW; B445B3638B8BEBE7 CRC64;

```

```

Query Match 97.9%; Score 1305; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 1.2e-79;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
QY 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATAEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTVVIEELKTIKIRSLRDEIGQLSLSKTLTSQ 240
QY 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273
Db 241 IALQKESDLCQSQIRETLSPRKSPSPSTKSS 273

```

```

RESULT 12
Q9AMA8
ID Q9AMA8 PRELIMINARY; PRT; 273 AA.
AC Q9AMA8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-12;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
RT trachomatis isolates with the Inca I47T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327005; AAG61102.1; -.
SQ SEQUENCE 273 AA; 30387 MW; 6945AE7E9B8AEBF3 CRC64;

```

```

Query Match 97.9%; Score 1305; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 1.2e-79;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGFL 60

```


Search completed: August 13, 2002, 09:30:17
Job time: 373 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:22:24 ; Search time 29.79 Seconds
(without alignments)
1145.073 Million cell updates/sec

Title: US-09-673-763-8
Perfect score: 1733
Sequence: 1 MTVSTNTSPVISRASPTF.....DLSAQPDENQSDAGEHKDS 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	2 S61491	inclusion membrane
2	200	11.5	390	2 A72108	hypothetical prote
3	200	11.5	390	2 B85114	hypothetical prote
4	185	10.7	276	2 H81707	inclusion membrane
5	164.5	9.5	273	2 C71553	probable inclusion
6	149	8.6	533	2 G72593	hypothetical prote
7	145.5	8.4	1732	2 T14039	protein kinase (EC
8	145	8.4	1690	2 T13030	microtubule bindin
9	143	8.3	1410	1 A57013	spindle pole body
10	141	8.1	944	2 S26710	outer membrane pro
11	140.5	8.1	403	2 H72216	methyl-accepting c
12	140	8.1	770	2 F82383	kinesin-like spind
13	140	8.1	1056	1 G02157	chromosome segrega
14	140	8.1	1155	2 A2675	structural mainten
15	140	8.1	1165	2 A97457	methyl-accepting c
16	139.5	8.0	739	2 H75001	myosin II heavy ch
17	139.5	8.0	746	2 T47237	hypothetical prote
18	139	8.0	1805	1 A64224	hypothetical prote
19	138.5	8.0	772	2 T27907	cytron - mouse
20	138.5	8.0	1597	2 S68420	skeletal myosin -
21	138.5	8.0	1957	2 A39294	paramyosin - fluke
22	137	7.9	439	2 A60608	hypothetical prote
23	136.5	7.9	1133	2 T22976	myosin heavy chain
24	136.5	7.9	1940	1 S04090	myosin heavy chain
25	136.5	7.9	1940	1 A24922	male-enhanced anti
26	136	7.8	1325	2 T42722	apsB protein - Eme
27	135.5	7.8	1051	2 T18302	myosin heavy chain
28	135.5	7.8	1937	2 I38055	interaptin - slime
29	135	7.8	1738	2 T14867	

30	134.5	7.8	628	2 F84219	Htr16 transducer I
31	134	7.7	1790	2 S67593	transport protein
32	133.5	7.7	284	2 C64527	M protein - Helico
33	133	7.7	539	2 F72288	methyl-accepting c
34	133	7.7	880	2 F75103	conserved hypothet
35	132.5	7.7	1475	2 T33318	hypothetical protei
36	132	7.6	773	2 T44989	hypothetical coile
37	132	7.6	1957	2 T38077	transducer protein
38	131.5	7.6	423	2 T44258	probable chemotaxi
39	131.5	7.6	739	2 F71161	synaptonemal compl
40	131.5	7.6	993	2 S49461	probable nuclear p
41	131.5	7.6	1837	2 T41023	myosin heavy chain
42	131.5	7.6	1999	1 S21801	hypothetical prote
43	131	7.6	327	2 F71980	methyl-accepting c
44	131	7.6	530	2 C72291	methyl-accepting c
45	131	7.6	569	2 D96954	methyl-accepting c

ALIGNMENTS

RESULT 1

S61491
inclusion membrane protein A - Chlamydomophila psittaci
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 31-Mar-2000
C:Accession: S61491
R:Mockey, D.D.; Heinzen, R.A.; Hackstadt, T.
Mol. Microbiol. 15, 617-626, 1995
A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a prote
A:Reference number: S61491; MUID:95302975
A:Accession: S61491
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <ROC>
A:Cross-references: EMBL:L35036; NID:9516598; PID:AA041443.1; PID:9516599

Query Match 100.0%; Score 1733; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.3e-93;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTVSTNTSPVISRASPTFGDHGKDFDNKIIPIISIEAPTSSAAAVCAKTAIEPEGKSP	60
DB	1	MTVSTNTSPVISRASPTFGDHGKDFDNKIIPIISIEAPTSSAAAVCAKTAIEPEGKSP	60
QY	61	LLQRCYLVKIIAATAIAFWGIAALVCLYLGSVISTPSLIILMLVSVFVITAIRDG	120
DB	61	LLQRCYLVKIIAATAIAFWGIAALVCLYLGSVISTPSLIILMLVSVFVITAIRDG	120
QY	121	TPSQVVRHMKQIQOFGENTRLHTAVENKAVNVELSEQINQLKHLTRLSDFGDRLEA	180
DB	121	TPSQVVRHMKQIQOFGENTRLHTAVENKAVNVELSEQINQLKHLTRLSDFGDRLEA	180
QY	181	NTGFTALIAIDFQLSLEEFKSVGKVTMTLSPFEKLAQSKETESQAVQAMMSVTELR	240
DB	181	NTGFTALIAIDFQLSLEEFKSVGKVTMTLSPFEKLAQSKETESQAVQAMMSVTELR	240
QY	241	TNLNALKELITENTVTEIQKADAEQVRFLEKRRQEEACSTLSHSIATLQESTT	300
DB	241	TNLNALKELITENTVTEIQKADAEQVRFLEKRRQEEACSTLSHSIATLQESTT	300
QY	301	LLKDSNTNLHVESRLIGVMVQDGAESTVVEASQDSDSAQPDENQSDAGEHKDS	355
DB	301	LLKDSNTNLHVESRLIGVMVQDGAESTVVEASQDSDSAQPDENQSDAGEHKDS	355

RESULT 2

A72108
hypothetical protein CP0581 [imported] - Chlamydomophila pneumoniae (strains CWL029 and
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: A72108; B81560

QY	16	SSPTFGDHGDFDNKKIIPISIEAP--TSSAAAVGAKTAIEPEGRSPLLQRIQCYLVKIIA	73
Db	2	TSPLVE-----MPLSCYPHTTSSTACTKRSSSLYKPSLIETQVRVAAFSL--	49
QY	74	AIALFVVGIAALVCLYIGSVITSTSLIIMLAILMVLFSVIVITADGTPTSPVRHMKQOI	133
Db	50	-ALLSIIGFLAILGHAIGFLIAPQVALVLIQAVFIISLGNALYICTKAPLRLYKELQDEV	101
QY	134	QQFGSEENTRLHTAVENLKAVNVELSEQINQLKOLHTRLSDFGDRLEANTGFTALIADFQ	191
Db	109	ASLKEVNFL-----LKSQKEF---LGLSKDFATTSKDLSD---VSLDFHNLQDFQ	153
QY	194	LSLSEFKSVGKTVMELSPFEKLAOSLKETTSQEAQVAMMSVTELRINI NAKEL----	241

Db 155 SSHQGF-----EDLEDYKNSAEDLRQIFSOETVQSLSKSTILSKKEIKVLPTEEV 207
 QY 250 --ITENKTVIEQLKADQA--LREQVRFLEKRKQOEACSTLSHSTAT--LOESTTLLKDS 305
 Db 208 RLRAENKEDLLKIVQLODIRK----LRAEINNLSQASKTILSEQIASQIENKLYANI 263
 QY 306 TTNL 309
 Db 264 TRAL 267

RESULT 5
 C71553
 Probable inclusion membrane protein A - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, A.; Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 Reference number: A71570; MUID:99000809
 Accession: C71553
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-273 <ARN>
 A:Cross-references: GB:AE001286; GB:AE001273; NID:g3328516; PIDN:AAC67710.1; PID:g332851
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: incA

Query Match 9.5%; Score 164.5; DB 2; Length 273;
 Best Local Similarity 22.8%; Pred. No. 0.011;
 Matches 67; Conservative 59; Mismatches 109; Indels 59; Gaps 10;
 QY 32 IIPISIEAPITSSAAVGAKTAEPEGRSPLLQRCILVLYKIAIALVFGIALVLYLG 91
 Db 8 VTPSPAPSYSANRV-----PPSLMDKIKKIAAASLTIGTIGFLALLGLHVG 58
 QY 92 SVISTPSLIIMLAIMVSVFVITAIRDTPSQVVR----HMKQOIQOFGCEENTRLHTAV 147
 Db 59 FLIAPQITIVLLALFIISLA-----GNALYLQKTANLHLQDQOR-----EV 100
 QY 148 ENLKAVNVLSSEIOLKQLHTRLSDFGDRLEANTGDTALIAADFOLSLSEEFKSVGTKE 207
 Db 101 GSLKEINFMLS--VLOKEFLH-----LSKEFATTSKDLASVSDQDFYSCLOGFRDNKGE 153
 QY 208 TMLSPFEKLAQSLKETFSQBAVQAMSSVTELRITNLNALKLITENK-----TVI 257
 Db 154 SLLDEYKNSTPEMKLFQSOEIIADLKGVSASLREERFLTPAEVRRRAHNOQSITVVI 213
 QY 258 EQLKA-DAQLRE--QVRFLEK-----RKQELSEACSTLSHSTATQESTT 300
 Db 214 BELKTRDSURDEIGLSQLSKTLTSQIALQRKESDLCQIRETSLSPRKSAS 267

RESULT 6
 G72593
 hypothetical protein APE1216 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 Reference number: A72450; MUID:99310339
 Accession: G72593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-533 <KAW>
 A:Cross-references: DDBJ:AP0000061; NID:g5104821; PIDN:BAA80205.1; PID:g1043991; PID:g510

A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1216

Query Match 8.6%; Score 149; DB 2; Length 533;
 Best Local Similarity 24.0%; Pred. No. 0.2; Mismatches 46; Indels 32; Gaps 7;
 Matches 59; Conservative 59; Mismatches 46; Indels 32; Gaps 7;

QY 124 QVVRHMKQOIQOFGCEENTRLHTAVENLKAVNVLSSEIOLKQLHTRLSDFGDRLEANTG 183
 Db 251 QAVEDLGLALDSLEERVGDLEFAVEDL---TLQLSSLDSSRVGALEDRVADIEGRLEAVEG 307
 QY 184 DF-----TALIAADFOLSLSEEFKSVGTKVTMLSPFEKLAQSLKETFSQ--EAVQAMMS 234
 Db 308 SLEDLSGAVDMSOOLQALAEDELSLSRVEDL---EARGVSVEDRLSQAEDIDSLTT 363
 QY 235 SVTELRITNLNALKLITENKTVIPOLKADQOLREEQVRFLEKRKQOELEACSTLSHSTAT 294
 Db 364 SLDSLRTELEDLSRLAEQAQASLEDNLTRLDVASTLQQLQRLATAEESLQALTEDLAS 423
 QY 295 LOESTTLLKDSITNLHAVESRLIGVMVQDGAESTVE-----EASODDSAQPDENQSD 348
 Db 424 LOAEVETLQOSIVE---IDRRLL-----CQLRSTVDVAVRLEVESLGEKLVQAEKNQRQ 473
 QY 349 AGEHKD 354
 Db 474 DASIED 479

RESULT 7
 T14039
 protein kinase (EC 2.7.1.37), myotonic dystrophy-associated - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14039
 R:Leung, T.; Chen, X.Q.; Tan, I.; Manser, E.; Lim, L.
 Mol. Cell. Biol. 18, 130-140, 1998
 A:Title: Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector
 A:Reference number: Z17862; MUID:98078670
 A:Accession: T14039
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-1732 <LEU>
 A:Cross-references: EMBL:AF021935; NID:g2736150; PID:g2736151; PIDN:AAC02941.1
 C:Genetics:
 A:Gene: MKR6
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hem
 C:Keywords: ATP; phosphotransferase
 F:75-343/Domain: protein kinase homology <KIN>
 F:1013-1062/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 8.4%; Score 145.5; DB 2; Length 1732;
 Best Local Similarity 21.1%; Pred. No. 1.3;
 Matches 54; Conservative 59; Mismatches 114; Indels 29; Gaps 7;

QY 126 VRHMKQOIQOFGCEENTRLHTAVENLKAVNVLSSEIOLKQLHTRLSDFGDRLEANTGDF 185
 Db 506 VNHLEQQLEANSVRRELDADFQIKAKFEKQIKTLOQEREELNKLQVQASERLNQSKEL 565
 QY 186 TALIAADFOLSLSEEFKSVGTKVTMLSPFEKLAQSLKETFSQBAVQAMSSVTELRITNL-- 243
 Db 566 KDAHQRKLAQMEFMEINERITELTQKQKLARVRO--KEEVEDLVNQKAESLRQLR 623
 QY 244 --NALKELITENKTVIPOLKADQOLREEQVRFLEKRKQOELE---EACSTLSHSTATL--- 295
 Db 624 AERAKKEVHTEALIAEASKDRKLREQSRHYSKOLENEGLEKQKQOISYSPGICSIHQ 683
 QY 296 QESTTLLKD-----STTNLHAVESRLIGVMVQDGAESTVE--GAESTVEE--ASODDS 338
 Db 684 QETIKLATDLEKKSIFVEEISRKREGIHAEINLKKELHDSGGQALNKLINVLKDKL 743

QY 339 AQPDENQSDAGEBKD 354
Db 744 EKTRRESQSEREFEN 759
RESULT 8
T13030
microtubule binding protein D-Clip-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul
A:Reference number: Z17588; MUID:98139549
A:Accession: T13030
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AA96783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton
Query Match 8.4%; Score 145; DB 2; Length 1690;
Best Local Similarity 20.6%; Pred. No. 1.3;
Matches 64; Conservative 57; Mismatches 99; Indels 90; Gaps 10;
QY 126 VRHMKQOQOGEENTRLTAV-----ENLKAVNVEL-----SEQINOLKOL 167
Db 1057 IKNQEEVTKAKTENLESTGTQTIKDLQERLEITNAELQHKEMASDAQIADLKT 1116
QY 168 -----HTRLSDFGRLEANTGDTALADFQSL-----EE 198
Db 1117 VEAIOVANANISATNAELSTVLEVLQAEKSETNHIFELFEMEADNMSRLIEKVTGKEE 1176
QY 199 FKSQVTKVETMLSPFEKLAQSLKE-----TFQBAVQAMSSVTEL 239
Db 1177 LKETHLQDEROKKFEELKLAQAEQKLOEQSQTSEKLTETIQOSLQELQDSVKOK 1236
QY 240 RTNLNALKELITENKTIVIE-----QLKDAQ--LREQVFLFKQOE--LEEAC 285
Db 1237 EELVQNLEEKVRESSIIIEAQNATKLNESNVQLENKTSCKETQDQLLESQKKEQLGEEA 1296
QY 286 STLSHSIATLOESTTLTKDSTTNLH-----AVESRLIGVMVQDGAESVTEE-----ASQ 335
Db 1297 AKLSGELQVQVQANGDIKDSLVKVEELVKVLEEKLOAATSQDQAQATNKELQLLVKSQ 1356
QY 336 DDAQAPQDEN 345
Db 1357 ENEGNLOGES 1366
RESULT 9
A57013
early endosome antigen 1 - human
N:Alternate names: endosome-associated protein
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A57013; S44243
R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell,
J. Biol. Chem. 270, 13503-13511, 1995
A:Title: EEA1, an early endosome-associated protein. EEA1 is a conserved alpha-helical p
A:Reference number: A57013; MUID:95286647
A:Accession: A57013
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1410 <RES>
A:Cross-references: GB:L40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368
R:Seellig, H.P.
submitted to the EMBL Data Library, April 1994
A:Reference number: S44243
A:Accession: S44243
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-254, 'C', 256-257, 'LO', 260-276, 'A', 278-283, 'A', 285-519, 'D', 521-574, 'BO', 5
A:Cross-references: EMBL:X78998; NID:g475933; PIDN:CAA55632.1; PID:g475934
C:Genetics:
A:Gene: GDB:EEA1
A:Cross-references: GDB:I369996
A:Superfamily: human early endosome antigen 1
C:Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane prote
Query Match 8.3%; Score 143; DB 1; Length 1410;
Best Local Similarity 24.0%; Pred. No. 1.4;
Matches 64; Conservative 52; Mismatches 99; Indels 52; Gaps 11;
QY 129 MKQOIQOQFEE-----NTRLHTAVENLKAVNVSELSQINQLKQLHTRLSDFGDRLEA 180
Db 812 LKODFETLSQETKIOHEELNNRIQTIVTELQVKMEKEALMTLSVVKDKLSKVSDSLKN 871
QY 181 NTGDF-----TALIADFQSLSEEFK-SVGTQVETMLSPFEKLAQSL-KETFSQEAQOA 231
Db 872 SKSEFEKENQKGAAILDLEKTKELKHQLOVQMENTLKEQKELKKSLEKEKEASHQLKL 931
QY 232 MMSVTE-----LRTNLNALKELITENKTIVIEQLKADAQALREEQVRF 273
Db 932 ELNMQEOLIOAONTLKQNEKEEQLOGNINELKQSEKQKQIEALQELKTIATLQKTE 991
QY 274 LE-KRQOELEEACSTLS---HSIATLQ-----ESTTLTKDSTTNLHVESRLIGVMVQDGA 325
Db 992 LENKLOOLTQAQALAAEKEKISVLQNNYKESQETFKQIQSDFYGRESELLATR-----Q 1047
QY 326 ESSTVEE---ASQDD--SAQPDENQS 347
Db 1048 DLKSVEEKLQAEEDLISNRNIGNQN 1074
RESULT 10
S26710
spindle pole body protein NUF1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D9476.3; protein YDR356W
C:Species: Saccharomyces cerevisiae
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 23-Mar-2001
C:Accession: S26710; S34288; A49455; S61152
R:Mirzayan, C.; Copeland, C.S.; Snyder, M.
J. Cell Biol. 116, 1319-1332, 1992
A:Title: The NUF1 gene encodes an essential coiled-coil related protein that is a pot
A:Reference number: S26710; MUID:92176232
A:Accession: S26710
A:Molecule type: DNA
A:Residues: 1-944 <MIR>
A:Cross-references: EMBL:Z11582; NID:g4069; PIDN:CAA77668.1; PID:g4070
A:Note: the authors translated the codon GAG for residue 206 as Asp and CTG for resid
R:Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
submitted to the EMBL Data Library, June 1993
A:Description: A spacer element in the Saccharomyces cerevisiae spindle pole body who
A:Reference number: S34288
A:Accession: S34288
A:Molecule type: DNA
A:Residues: 1-944 <KIL>
A:Cross-references: EMBL:X73297; NID:g312173; PID:g312175
R:Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
J. Cell Biol. 123, 1175-1184, 1993
A:Title: A spacer protein in the Saccharomyces cerevisiae spindle pole body whose tra
A:Reference number: A49455; MUID:94064779
A:Accession: A49455
A:Molecule type: DNA
A:Residues: 1-22 <K12>
A:Cross-references: EMBL:X73297
R:Du, Z.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of S. cerevisiae cosmid 9476.

A;Reference number: S61148
A;Accession: S61152
A;Molecule type: DNA
A;Residues: 1-944 <DU2>
A;Cross-references: EMBL:U7
C;Genetics:
A;Gene: SGD:NUF1; SPC110
A;Cross-references: SGD:S00
A;Map position: 4R
C;Keywords: coiled coil; m

Query Match 8.1%; Score 141; DB 2; Length 944;
Best Local Similarity 23.5%; Pred. NO. 1.1;
Matches 61; Conservative 56; Mismatches 101; Indels 42; Gaps 8;

	Matches	61; Conservative	56; Mismatches	101; Indels	42; Gaps
Qy	129	MKQIQOFGENTRLHRTAVENLK-----AVNVFELSEQINQLKOLHTRLRSLDFGDR	177		
		: : : : : : : : : : : : : : : : : : :			
Db	316	MDLQLKQKQKESKRLKDELNELETKFSNGSQSSAKENELKMLKNKIAELEEEISTKNSQ	375		
		: : : : : : : : : : : : : : : : : : :			
Qy	178	LEANTGPTALIAFDQISLEEFK-----SYGTVETMLSPFEKLAQSLK---ETESQ--	226		
		: : : : : : : : : : : : : : : : : : :			
Db	376	LIAKEGLASLMA--QLTQLESKLNQRDSQLGSREBELKKTNDLQKDIRAREETVSKD	433		
		: : : : : : : : : : : : : : : : : : :			
Qy	227	EAVQAMSSVTELTNLNALKELITENKTV-----IEOLKADAQALUREEQVRFLE	275		
		: : : : : : : : : : : : : : : : : : :			
Db	434	ERIIDLQKKVKQLENDLFVTKTHSESKTTIDNLESKDKLIKILENDLKAQEKYSKME	493		
		: : : : : : : : : : : : : : : : : : :			
Qy	276	KRKQELBEAGSTLHSHSTATIQESTTLKDKSTTNLHAYVESRLIGVWGDGAESSIVEASQ	335		
		: : : : : : : : : : : : : : : : : : :			
Db	494	K---ELKREFNYKISSEKLEDEKTTLINEKISLNLAENSQLNKKI-----EDNSTATHMK	546		
		: : : : : : : : : : : : : : : : : : :			
Qy	336	DDSAQPDQENQSDAGEHKDS	355		
		: : : : : : : : : : : : : : : : : : :			
Db	547	ENYEKOLESRLKDIIEYKES	566		
		: : : : : : : : : : : : : : : : : : :			

RESULT 11

H72216
outer membrane protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72216
R:Nelson, M.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
Accession: H72216
Status: preliminary
Molecule type: DNA
A:Residues: 1-403 <ARN>
A:Cross-references: GB:AE001812; GB:AE000512; NID:g4982302; PID:g4982300
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1729

Query Match

Best Local Similarity	21.08;	Score 140.5;	DB 2;	Length 403;
Matches 73;	Conservative	75;	Mismatches 113;	Indels 87;
Gaps	15;			

QY	47	VGAKTAIEPEGRSP	LQRICYLKIIIAAIA	LVFGVIAALVCLY	IGSVISTPSLILMAIM	106
Db	40	LGVLTM-PDG----	TFQGNSLTRYQRAVAFY	-----	RLYNILKOPSDVSGLIN	85
QY	107	LVSTV---IVITAIR	-----	DGTPSQV-----	VRHMK-----	QIQQ 135
Db	86	KVSTLEDLVSTATMK	VQNLSDNFGGVTS	SDLETKLVNDVANL	KATLVDLKLVREVM	SQVS 145
QY	136	FGEE-----NTRL	HTAVENLKA	VNVELSEQI-----	NQLKQLHRLSDFDRL	---EAN 181

Db	146	QSDQLQSDLADKVNKAESKLSKSGDFVNKDYVDSKIAQTQVSKLSLDSLEGRLSAVETK	205
QY	182	TGDETALIAFDQLSLEEFKSVGKVTMLSPFEKLAQSLKETFSQAVQAMWSVTLEKT	241
Db	206	TANLEALVRNSEASKDY-----VBKTLKSYTDTLQDQLKLSLSAS-----VBK	248
QY	242	NLNALKELTENTVTIBQLKADALREEQVRFLEKRRQBLEEACSLSHSIATLQBSTTL	301
Db	249	NNTALSGEIGNLVKSLQSDLETQOKTARALDARVSLVGQITTVNSRVSELEKRVSQ	308
QY	302	LKDSITTLNLHAVESRLICVMQDGAESTVEEASQDSDSAPODENQSDA	349
Db	309	VEGADVKNLSLE-RSMGAVT---ARYTKVEE-----EVKNLNQNSNA	345

RESULT 12

F82383
methyl-accepting chemotaxis protein VCA1056 [Imported] - Vibrio cholerae (strain N1659)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82383
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: F82383
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <HEI>
A:Cross-references: GB:AE004432; GB:AE003853; NID:g9658509; PIDN:AAF96950.1; GSPDB:EN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA1056
A:Map position: 2

Query Match

Query Match	8.1%;	Score 140;	DB 2;	Length 770;
Best Local Similarity	21.6%;	Pred. No. 1;		
Matches	73;	Conservative	60;	Mismatches 139; Indels 66; Gaps 13;
Qy	68	LVKIIAALFWGIAALVCLYLSVISTP--SLILMLAIMLVFVIVITAIRGTP----	122	
Db	417	IMKVLLAIVGAV-IATLAAMVWGISKIPRDSIQIRMSRDNDITVRLSEQSGSDEIR	475	
Qy	123	-----SOVVRHMKOOIQQFGE-----NTRLHTAVENLKAHVNLSEQ-----IN	162	
Db	476	QLAQAALRMLSHLQDTICQFAHTTDKLNHSHTQTITHNMTGRNSVSQHERHDSVVTAVN	535	
Qy	163	OLKQLHRLSLDFGDRL-----EANTGDETLADIAPQLSLEEFKSVGPKVTMLSPFEKLA	217	
Db	536	EMASITFSEVFEAQRATFVQEAENKQHGGSVGNELA-RDMTISIN---OOMASAVEAIA	591	
Qy	218	QSLKETFSQEAQVAMMSVTELRTNLAKELTIENTK-----VIGOLKADAQ	265	
Db	592	RLNHESQSIASVLDVIOQIAE-QTNLLALNAATEARGQGRGFVAVADEVRNLAAKTQ	650	
Qy	266	LREQVRP-----LEKRQE-----LEEACSTLSHSIATPLQESTTLLKDDSTTNLHAVESRLI	317	
Db	651	TSTEEIRTKIDRLQKETQSVVNVCEEAENNVVVRGATCSHTNDMLKQIVDNLNLNELNMNI	710	
Qy	318	-----GVMVQDGAESTVVEEASQDDSAQPDEN	345	
Db	711	QIATATEQQRGVTEENANITISDSVSYASVTVQGVGN	748	

RESULT 13

G02157
kinesin-like spindle protein HKSP - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: G02157

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:21:19 ; Search time 53.52 seconds
(without alignments)
736.756 Million cell updates/sec

Title: US-09-673-763-8

Perfect score: 1733

Sequence: 1 MTVSTDNTPVISRASPTE.....DDSAQPQDENSDAGEHKDS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	20 AAY32173	Chlamydia psittaci
2	200	11.5	397	20 AAY34783	Chlamydia pneumoniae
3	162.5	9.4	273	20 AAY32176	Chlamydia trachoma
4	149	8.6	612	22 AAB95546	Human protein sequ
5	145	8.4	1690	22 ABB61144	Drosophila melanog
6	145	8.4	1690	22 ABB61173	Drosophila melanog
7	143.5	8.3	1286	21 AAB43359	Human ORF3123
8	143.5	8.3	2053	22 AAB43359	Human protein kina
9	143	8.3	1411	17 AAW02258	Nucleolar/endosoma
10	140.5	8.1	434	21 AAB42352	Human ORF ORF116
11	140	8.1	575	22 AAB47215	Human KSP-S553. S

12	140	8.1	1057	22 AAG67419	Amino acid sequenc
13	140	8.1	1057	22 AAB47212	Human KSP. Homo s
14	139.5	8.0	739	22 AAB96493	Putative sensory t
15	138.5	8.0	344	21 AAB42170	Human ORF ORF1934
16	136	7.8	1325	18 AAW19540	Male-enhanced anti
17	136	7.8	1325	20 AAW94391	Mouse male enhance
18	134	7.7	475	22 ABB68543	Drosophila melanog
19	133.5	7.7	1948	22 ABB68543	Novel human diagno
20	133	7.7	880	22 AAB96332	Putative P. abyss
21	131	7.6	286	18 AAW20404	H. pylori secreted
22	131	7.6	286	18 AAW24652	H. pylori secreted
23	131	7.6	327	18 AAW20901	H. pylori secreted
24	131	7.6	2482	16 AAW2826	Human mitotin amin
25	131	7.6	2482	19 AAW23996	Human mitotin amin
26	131	7.6	3248	17 AAR99795	Kinetochore protel
27	130.5	7.5	945	20 AAW73624	Human secreted pro
28	130.5	7.5	1521	21 AAG39235	Arabidopsis thalia
29	130.5	7.5	1528	21 AAG39234	Arabidopsis thalia
30	130.5	7.5	1562	21 AAG39233	Arabidopsis thalia
31	130.5	7.5	1703	21 AAG36714	Arabidopsis thalia
32	130.5	7.5	1710	21 AAG36713	Arabidopsis thalia
33	130.5	7.5	1728	21 AAB42089	Arabidopsis thalia
34	130.5	7.5	1744	21 AAB42089	Arabidopsis thalia
35	130.5	7.5	1797	21 AAG36712	Arabidopsis thalia
36	130.5	7.5	1797	22 ABB13880	Novel human diagno
37	130	7.5	2779	22 ABB62371	Drosophila melanog
38	130	7.5	1184	22 AAG33733	Staphylococcus aur
39	129.5	7.5	1188	22 AAG36532	Staphylococcus aur
40	129.5	7.5	1483	22 AAG05140	Novel human diagno
41	129.5	7.5	1493	21 AAG48639	Arabidopsis thalia
42	129.5	7.5	1544	21 AAG48638	Arabidopsis thalia
43	129.5	7.5	2633	22 AAG06505	Arabidopsis thalia
44	129	7.4	1553	22 AAG02200	Novel human diagno
45	129	7.4	1788	22 AAG06749	Novel human diagno

ALIGNMENTS

RESULT 1
AAY32173
ID AAY32173 standard; Protein; 355 AA.
XX AC
XX AAY32173;
XX AC
XX 01-FEB-2000 (first entry)
XX DT
XX DE Chlamydia psittaci infection-specific protein Inca.
XX Inca; Infection; vaccine; therapy; diagnosis.
XX OS Chlamydia psittaci.
XX PN WO9953948-A1.
XX PD 28-OCT-1999.
XX PF 20-APR-1999; 99WO-US08744.
XX PR 20-APR-1998; 98US-0082438.
XX PR 21-APR-1998; 98US-0082588.
XX PR 22-MAY-1998; 98US-0086450.
XX (UYOR-) UNIV OREGON STATE.
XX Rockey DD, Bannantine JP;
XX WPI; 1999-633904/54.
XX N-PSDB; AAZ34587.
XX Novel bacterial infection specific proteins for treating and diagnosing
XX chlamydial infections

XX
Griffais R:

WPI; 1999-357842/30.
Genome sequence of *Chlamydia pneumoniae*
Accession: 1912bp; English.

Sequence 355 AA; 100 % Score 1733. DB 20: Length 355;

CC especially where the vector infects the caprine
CC epitope of *C. pneumoniae*.
XX
XX Sequence 397 AA;
SQ

Query Match	11.5%;	Score 200;	DB 20;	Length 397;
Best local similarity	23.0%;	Pred. No. 4.7e-08;		
Matches	87; Conservative	64; Mismatches	142; Indels	86; Gaps
13:				
Qy	8	TSPVTSRASSPTFGDHGDKFDNNKIPISTIEAPTSSAAAVGAKTAIEPGRPQLQRIC	67	
Db	9	spsvntptsapn-----tpipatp---giptt---kprsfiekvi-	45	
Qy	68	LVKIIIAATALFVVGAALVCLYLGVSIS-----TPSL-LIMIAMLVSVVITAIRDGT	121	
Db	46	--ivakylilfai--aatsgaltilgslagtpigialliffvsmvilgliikdsi	99	
Qy	122	PSQVVRHMKOOITOEGEEN-----TRLHTAVENLKAVN-----	154	
Db	100	sggeerlrfeevsrfsenqriltvittletevkdkaaqdgltleiafrnengnlktt	159	
Qy	155	-VELSQINOLKQHLRLSDFGRLEANTGDFTALIADQLSLSEFKS-VGTVEVTMLSP	212	
Db	160	aedleeqvaklsesqealerinqldiganagdaqselselklisgwskvveqintsi--	217	
Qy	213	FELAQSLKETFSQEAVQAMMSVTELRNLNLKELI-----TENKTVIQOLKADAQ	265	
Db	218	----qalkvllogewvveaqthvakwediqalgaelgmhgngstalgksvenllvqdq	272	
Qy	266	LREEQVFLEKKRQOELEEACSTLSHSSTATLQESTTLLKDSTTNLHAVESRLIGVMVDGA	325	
Db	273	altrrvvgellesenkisqacsalrqeleklahetslqqridamladeqn-----a	324	
Qy	326	ESSTVEASODDSAAQPQDE	344	
Db	325	egvtalekmqeaakaese	343	

ULT 2
34783
34783 standard: protein: 397 AA:

88534703.

AAI 34 / 83,

13-SEP-1999 (first entry)

Chlamydia pneumoniae transmembrane

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

Respiratory disease; pneumonia; bronchitis; purulent otitis media; emphysema

vaccine: neutralising epitope.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Chlamydia pneumoniae.

WO9927105-A2.

02 - JUN - 1999

03 JUN 1955.

20-NOV-1998; 98WO-IB01890.

04 - NOV-1998;
98US-0107078.

21-NOV-1997; 9/FK-0014673.

(CFST) GENSET

FLOWER (1976)

Tue Aug 13 14:23:42 2002

```

Db 245 lknemenerwhlgtkietklqkemadiveasrtstlelqnlqdeykeknrelaemgrqlik 304
Qy 304 DSTNLIHAVESRLIGVMQDGAESSTVEPASODSQAQODE 344
Db 305 ekt--leaeksritamkmgdmr--lmeelrdrqagde 340

RESULT 5
ABB61144
ID ABB61144 standard; Protein; 1690 AA.
XX
AC ABB61144;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10224.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR N-PSDB; ABL05247.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 10224; 21pp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1690 AA;

Query Match 8.4%; Score 145; DB 22; Length 1690;
Best Local Similarity 20.6%; Pred. No. 0.011;
Matches 64; Conservative 57; Mismatches 99; Indels 90; Gaps 10;

QY 126 VRHMKQIQQFGENTRLHTAV-----ENLKAVNVNEL-----SEQINOLKOL 167
Db 1057 lknleqvktktenlelsgtqttdlgerlelthaelqhkekmasedaqkiadlktl 1116
QY 168 -----HTRLSDFCDRLEANTGDFTLADIADQLSL-----EE 198
Db 1117 veaigvanisanatnaelstvlvlelqaeksetnhlfefemadmsrlliekvtgikee 1176
QY 199 FKSvGTkvETMLSPFEKLAQSLKE-----TFSPQAVAMMSSVTEL 239

```

```

Db 1177 lkehlqldrqkffeelkekklqagseqklqesqtsketeiqgslqelqdvkqk 1236
Qy 240 RTNLNALKELITENKTVE-----OLKADAO-LREEQVRFLEKKEQ--LEEAC 285
Db 1237 eelvqnleekvressiieaqntklnesnvqlenktclktqdllesqkkekqlqee 1296
Qy 286 STLSHSIATLOESTWLLKDDSTNLH---AVESRLIGVMQDGAESSTVEE-----ASQ 335
Db 1297 akisglqgvqangdikdsvkveelvkvleeklqaatsqldaqatnkelqellvksq 1356
Qy 336 DDSAQPDEN 345
Db 1357 enegnlgges 1366

RESULT 6
ABB61173
ID ABB61173 standard; Protein; 1690 AA.
XX
AC ABB61173;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10311.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR N-PSDB; ABL05276.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 10311; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1690 AA;

Query Match 8.4%; Score 145; DB 22; Length 1690;
Best Local Similarity 20.6%; Pred. No. 0.011;
Matches 64; Conservative 57; Mismatches 99; Indels 90; Gaps 10;

QY 126 VRHMKQIQQFGENTRLHTAV-----ENLKAVNVNEL-----SEQINOLKOL 167

```

Db 1057 iknlqeetvaktaktenlelstgtgttkldqlerleithnaelqhkemasedaqkiadkl 1116
QY 168 -----HTRLSPDGRLEANTGDTTALADQLSL-----EE 198
Db 1117 vealqvanaisatnaelstvlvldqeksethifelfeemadmnserliekvgtgikee 1176
QY 199 FKSQVTKVETMLSPFEKLAQSLKE-----TFSQEAVQAMMSSVTEL 239
Db 1177 lkethiqldqrqkfeeleeklkqagseqklqgesqtskektelelqglqelqdsvkqk 1236
QY 240 RTMLNALKELITENKTVIE-----OLKADAQ-LREQVRPLEKRRQE--LEEAC 285
Db 1237 eelqvleekvressiieaqtlnesnvlenktsciketqgdqlesqkqkqleq 1296
QY 286 STLHSHIATLOESTTLKDDSTNLH----AVESRLIGVMVQDGAESSTVEE-----ASQ 335
Db 1297 akisgelqvgqangdikdlskvveelkvleeklqaatsqldaqatnkelqellvksq 1356
QY 336 DQSAQPDQDEN 345
Db 1357 enegnlqges 1366
RESULT 7
AAB43359
ID AAB43359 standard; Protein; 1286 AA.
AC AAB43359;
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF3123 polypeptide sequence SEQ ID NO:6246.
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiporiatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX W0200058473-A2.
PD 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Leach M;
PI WPI; 2000-602362/57.
XX N-PSDB; AAC77568.
DR Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX Claim 11: Page 5433-5436; 5507pp; English.
PS

XX AAC74446 to AAC77606 encode the proteins given in AAB40337 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiporiatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1286 AA;
Query Match 8.3%; Score 143.5; DB 21; Length 1286;
Best Local Similarity 24.3%; Pred. No. 0.0099;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
QY 127 RHMKQ---IQQFGENTRLHTAVENLKAIVNELSQI-----NOLKQLHRLS 172
Db 74 rmnkqgemseirrqkfyetqagkleagrnkleeqkshqhsdknrilletrlr 133
QY 173 DFGDRLEANTGDTTALADQLSLSEEFKSVGTVKVTMLSPFE-KLAQS---LKETF--SQ 226
Db 134 evsleheeqklelkrqltelqlslqeresqqltalqaaraalesqirqrakteleetaae 193
QY 227 EAVQAMMSSVTELRYNLNALKKE---LITENKTVIFQLKAD----- 263
Db 194 eelqatahrdeqlrkdalrnsctvitdleeqlnqltednaelnqnfyiskqideasg 253
QY 264 -----AQLREEVFRLEKRRQKEEACSTLSHSIATLOESTTLKDDSTNLHVESRLI 317
Db 254 andeivqrse-vdhlrreiteremqltsqkqmealkttctmleeqvmdlealndell 311
RESULT 8
AAU03501
ID AAU03501 standard; Protein; 2053 AA.
XX AC AAU03501;
XX 12-SEP-2001 (first entry)
XX Human protein kinase #1.
DE Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder.
XX Homo sapiens.
OS
XX W0200138503-A2.
XX 31-MAY-2001.
PD 22-NOV-2000; 2000WO-US32085.
XX 24-NOV-1999; 99US-0167482.
XX (SUGE-) SUGEN INC.
PA

XX PF 27-APR-1995; 95DE-1015514.
XX PR 27-APR-1995; 95DE-1015514.
XX PA (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX PI Renz M. Seelig HP;
XX WPI; 1996-403153/41.
XX DR N-PSDB; AAT58751.
XX DNA encoding nucleolar-endosomal auto-antigen - useful for exact
PT diagnosis of rheumatic disease, in gene therapy and for removal of
PT specific auto-antibodies
XX Claim 1; Fig 2; 15pp; German.
XX Transformed cells can be cultured to produce the antigen p162, for use
CC in exact (differential) diagnosis of rheumatic disease, i.e. they
CC can detect, in immunoassays, Western blots, etc., rheumatism-
CC specific auto-antibodies. The antigen can be used therapeutically,
CC in the removal of auto-antibodies from the circulation, or when
CC coupled to a cytotoxin, the elimination of auto-antibody-
CC producing lymphocytes.
XX SQ Sequence 1411 AA;
Query Match 8.3%; Score 143; DB 17; Length 1411;
Best Local Similarity 24.0%; Pred. No. 0.012;
Matches 64; Conservative 52; Mismatches 99; Indels 52; Gaps 11;
QY 129 MKQIQOQFEE-----NTRLHTAVENLKAVNVSEIQLKOLHTRLSDFGDRLEA 180
Db 812 lkqdfetlsqetkiqheelnriqtvtelqkvmekealmteistvkdsksvsdln 871
QY 181 NTGDF-----TALIADFOLSEEFK-SVGTKVETMLSPFEKLAOSL-KETFSQAVOA 231
Db 872 sksefekenqkaalidlektkelkhqlqvqmentlkeqkelkslekeashqlk 931
QY 232 MMSSVTE-----LFTNLAKELITENKTVIEOLKADAOLREBQVRF 273
Db 932 elnsmqeqliqaqnlknekeeqqlqgninelkqsseqkqkqlealqgelkviakte 991
QY 274 LE-NRKELEFACSTLS---HSIATLQ-----ESTTLKDDSTTNLHAVESKLIGVMVDGA 325
Db 992 lenklqqqltqaacqlaakekiskvislqnnnyeksgqetfkqlqsdqfygresellatr-----q 1047
QY 326 ESSTVEE---ASQDD---SAQPQDENQS 347
Db 1048 dlksveeklsiqagedllsnrnqignqn 1074
RESULT 10
AAB42352
ID AAB42352 standard; Protein; 434 AA.
XX AC AAB42352;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2116 polypeptide sequence SEQ ID NO:4232.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX PD 12-SEP-1996.
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI Flanagan P, Clary D;
XX WPI; 2001-343950/36.
XX DR N-PSDB; AAS06701.
XX Nucleic acids encoding human kinase polypeptides, useful for preventing
PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and
PT neuronal-associated diseases, and microbial infections -
XX Claim 7; Figure 2; 433pp; English.
XX AAU03501-AAU03557 represent novel human protein kinases #1-57. The
CC novel protein kinases have been identified as members of the tyrosine
CC or serine/threonine kinase (PTK and STK) families. The polynucleotides
CC encoding protein kinases and the polypeptides may be used in the
CC prevention, diagnosis and treatment of diseases associated with
CC inappropriate kinase expression. For example, they may be used to treat
CC cancers (especially cancers of haematopoietic origin), cardiovascular
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
CC immune related diseases (e.g. rheumatoid arthritis), neurological
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
CC Additionally, polynucleotides encoding protein kinases may be
CC used for gene therapy and as DNA probes in diagnostic assays.
CC The protein kinase polypeptides may be used as antigens in the production
CC of antibodies against the protein kinases and in assays to identify
CC modulators of protein kinase expression and activity.
XX SQ Sequence 2053 AA;
Query Match 8.3%; Score 143.5; DB 22; Length 2053;
Best Local Similarity 24.3%; Pred. No. 0.019;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
QY 127 RHMKQO---IQFGEENTLHTAVENLKAVNVSEIQLKOLHTRLS 172
Db 843 rnmkaqeemiselrgqkfyelqagkileaqnkrkleekiekishqhdsknrilleltr 902
QY 173 DFGDRLEANTGDTALIADFOLSEEFKSVGTKVETMLSPFE-KLAQS---LKETF--SQ 226
Db 903 evsleheeqklelkrqltelqlslqeresqltalqaaraalesqlrqakteleetaeae 962
QY 227 EAVQAMMSSVTELTNLALKE---LITENKTVIEOLKAD----- 263
Db 963 eeiqaltahrdelqrkfdalrnsctvitdeeqinqitdnealnqnfylskqldeasg 1022
QY 264 ----AQLREEQVRLEKRRQLEECACSTLSHSIATLQESTTLKDDSTTNLHAVESRLI 317
Db 1023 andeivqlrse-vdhrirletermqtsqkqmealktctmleesqvmdealdell 1080
RESULT 9
AAW02258
ID AAW02258 standard; Protein; 1411 AA.
XX AC AAW02258;
XX DT 09-MAR-1997 (first entry)
XX DE Nucleolar/endosomal auto-antigen p162.
XX Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW gene therapy.
XX Homo sapiens.
OS DE19515514-C1.
PN DE19515514-C1.
XX PD 12-SEP-1996.

QY 113 VTAIRDGP-----SQVVRHMQQIQFGEENTRLHTAVENLKAVNTSELSQNQLKQ 166
||||| : || |:: : | : | : | :
pb 298 vitalvertpbhvpyresktrilqdslg-grtrtsi---iataspaslnleetstley 352

Query Match 8.1%; Score 140; DB 22; Length 1057;
Best Local Similarity 21.7%; Pred. No. 0.015;
Matches 56; Conservative 100; Indels 46

QY 113 VITAIKDTGP-----SOVVRHMKQIQOFGENTRLHTAVENLKAVNVSEIQNLQK 166
 Db 298 vitalvertphvreskltrilqdsig--grtrtsi---iatispaslnleettstley 352
 QY 167 LHTRLSDFGDRELEANTG--DFTALIDFQSLSEEFK----SVGTRKVTMLSPFEKLAQSLK 221
 Db 353 ah-raknllinkpevnqkltkkalikeyteelerikrdlaaarekngvyiseefrnmsgk 411
 QY 222 ETSQEAQVAMSSVTELRNLKALITENKTVIEOLKADQALREQVRFLKRRKQEL 281
 Db 412 ltvqeegivliekigaveeelrvrtelfmdnkneidqcksdignktqelctgkhlqet 471
 QY 282 EACSTLSHIAVLQES--TTLKLDSTNHLHAVESRLIGVMVDGAESSTVEASOD--- 336
 Db 472 k-----lqivkeeyitsalesteeklhdaaskil-----ntveettkdvsg 512
 QY 337 -----DSAQPDQENQSDA 349
 Db 513 lnskldrrkkavdqhnaea 530

RESULT 14
 ID AAB96493 standard; Protein; 739 AA.
 AC AAB96493;
 DT 29-OCT-2001 (first entry)
 DE Putative sensory transduction histidine kinase and response regulator #3.
 KW Hyperthermophilic archaeon; hyperthermophilic protein.
 OS Pyrococcus abyssi.
 PN FR2792651-A1.
 PD 27-OCT-2000.
 PF 21-APR-1999; 99FR-0005034.
 PR 21-APR-1999; 99FR-0005034.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Hellig R;
 PI WPI; 2001-126236/14.
 PS New nucleotide sequences isolated from Pyrococcus abyssi encode
 proteins useful in industry -
 Claim 7; Pages 1203-1205; 1657pp; French.

The present invention relates to the genomic sequence of Pyrococcus abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present invention is one such P. abyssi protein. The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to 110 degrees centigrade.
 Note: This patent is in the same patent family as WO2000065062, which contains additional sequences as shown in AAB99132-AAB99143, AAB75903-AAH75920 and AAG66436.
 Sequence 739 AA.

Query Match 8.0%; Score 139.5; DB 22; Length 739;
 Best Local Similarity 21.1%; Pred. No. 0.0096;
 Matches 63; Conservative 66; Mismatches 124; Indels 45; Gaps 13;

QY 71 IIAIAL-FVWGIAALVCLY-LGSVISTPSLILMLAILVSVFIVITAIKDTGPSQVVRH 128
 Db 308 viqgfavgfviilvavilviykiasntlap-----leklyaaqalaegrlkqvsey 358
 QY 129 MKQOIQOFGENTRLHTAVENLKAVNVSEIQNLKQLHTRLSDFGDRLEANT---GD 184
 Db 359 lk-qiryl--erdegaligafavskdlvgtlnaisklerlae-gdlsngltvevrge 414
 QY 185 FTALIDFQSLSEEFK--SVGTRKVTMLSPFEKLAQSLKETFQ--EAVQAMSSVTELR 241
 Db 415 lrdiellirsvtetfreslgsive-mandlekranlaqskdvtealngvnealqgvs 473
 QY 242 NLNALKEKITENKTVIEOLKADQALREQVRFLKRRKQELSEACS----- 286
 Db 474 eaqrqetline---itdgmrlvadtseesvrameefsgavtevvsvianegsqgdealkr 530
 QY 287 --TLSHSIATLQESTTLTKDSTNHLHAVESRLIGVMVDG--AESSTVEASQDSDAQ 340
 Db 531 lediqhmsrieetvskvaemsrnieitnvtisaeqtnllalnaaeaaaraagegr 588

RESULT 15
 ID AAB42170 standard; Protein; 344 AA.
 AC AAB42170;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1934 polypeptide sequence SEQ ID NO:3868.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 PN WO2000058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000WO-US08621.
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Leach M;
 PI WPI; 2000-602362/57.
 DR N-PSDB; AAC76379.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -
 PS Claim 11; Page 3015-3016; 5507pp; English.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:24:24 ; Search time 17.41 Seconds
(without alignments)
789.514 Million cell updates/sec

Title: US-09-673-763-8
Perfect score: 1733
Sequence: 1 MTVSTDNTSPVIRASSPTF.....DSSAQPDENQSDAGEHKDS 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.5	8.3	1286	1	CTRO_HUMAN
2	141	8.1	944	1	NUF1_YEAST
3	140	8.1	1057	1	EGS_HUMAN
4	139	8.0	866	1	MYSP_SCHMA
5	139	8.0	1805	1	HMW2_MYCGE
6	138.5	8.0	1597	1	CTRO_MOUSE
7	136.5	7.9	1940	1	MYH3_HUMAN
8	136.5	7.9	1940	1	MYH3_RAT
9	136	7.8	1325	1	GLI6_MOUSE
10	135.5	7.8	1937	1	MYH8_HUMAN
11	134	7.7	1790	1	USO1_YEAST
12	133.5	7.7	1939	1	MYH4_HUMAN
13	133	7.7	880	1	RA50_PVRAB
14	132	7.6	433	1	HTR2_HALVA
15	132	7.6	539	1	MY53_HYDAP
16	132	7.6	1957	1	YD86_SCHPO
17	131.5	7.6	516	1	P54_ENTFC
18	131.5	7.6	993	1	SCP1_MOUSE
19	131	7.6	1969	1	MYSA_CAEEL
20	131	7.6	3210	1	CENF_HUMAN
21	131	7.6	3911	1	AKA9_HUMAN
22	130	7.5	886	1	RA50_SULAC
23	129.5	7.5	2663	1	CENE_HUMAN
24	129	7.4	1972	1	MYHB_HUMAN
25	128.5	7.4	473	1	YVCE_BAGSU
26	128.5	7.4	1938	1	MYH4_RABIT
27	128.5	7.4	1939	1	MYH1_HUMAN
28	128	7.4	2022	1	ANT1_ONCVO
29	127.5	7.4	2349	1	TPR_HUMAN
30	127	7.3	1938	1	MYHD_HUMAN
31	127	7.3	1972	1	MYHB_MOUSE
32	127	7.3	1976	1	MYHA_RAT
33	126.5	7.3	1084	1	MYSS_RABIT
					P02562 oryctolagus

34	126.5	7.3	1130	1	YL17_CABEL	Q11102 caenorhabdi
35	126.5	7.3	1972	1	MYHB_RABIT	P35748 oryctolagus
36	126.5	7.3	1978	1	MYHB_CHICK	P10587 gallus gall
37	126	7.3	866	1	MYSP_SCHJA	Q05870 schistosoma
38	126	7.3	882	1	RA50_PVRFU	P58301 pyrococcus
39	126	7.3	1244	1	MIX1_CABEL	Q09591 caenorhabdi
40	126	7.3	1433	1	REST_CHICK	Q42184 gallus gall
41	126	7.3	1976	1	MYHA_HUMAN	P35580 homo sapien
42	126	7.3	2245	1	MYSU_DICDI	P54697 dictyosteli
43	125.5	7.2	1184	1	BIMC_EMEI	P17120 emericella
44	125.5	7.2	1966	1	MYSB_CAEEL	P02566 caenorhabdi
45	125	7.2	697	1	MFPI_LYCES	P93203 lycopersico

ALIGNMENTS

RESULT 1
CTRO_HUMAN
ID CTRO_HUMAN STANDARD; PRT; 1286 AA.
AC OI4578; Q9UPZ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Citron protein (Rho-interacting, serine/threonine kinase 21)
DE (Fragment).
GN CIT OR STK21 OR KIAA0949.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Connell M., Goela D., Harper M.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 347-1286 FROM N.A.
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
CC -!- FUNCTION: PUTATIVE RHO/RAC EFFECTOR THAT BINDS TO THE GTP-BOUND
CC FORMS OF RHO AND RAC1. IT PROBABLY BINDS P21 WITH A TIGHTER
CC SPECIFICITY IN VIVO (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AC002563; AB071327.1; -;
DR EMBL; AB023166; BAA76793.1; -;
DR MIM; 605629; -;
DR InterPro; IPR001180; CNH.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000861; REM_repeat.
DR Pfam; PF00780; CNH; 1.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00036; CNH; 1.
DR SMART; SM00233; PH; 1.

[illegible]

RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEOSKELETON. POTENTIAL
CC ROLE IN CROSSLINKING FILAMENTS OR ANCHORING OTHER MOLECULES. IT
CC IS ESSENTIAL FOR GROWTH.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; TIGHTLY ASSOCIATED WITH THE
CC NUCLEUS. IT IS PRESENT IN A GRANULAR PATTERN THAT EXCLUDES THE
CC NUCLEOLUS.
CC -!- PTM: MAY BE REGULATED BY PHOSPHORYLATION EVENTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 211582; CAAT7668.1; -
CC EMBL; X73297; CA51733.1; -
CC EMBL; U28372; A4664791.1; -
CC PIR; S26710; S26710.
CC PIR; S34288; S34288.
CC SGD; S0002764; NUF1.
CC Coiled coil; Nuclear protein; Phosphorylation.
CC FT DOMAIN 164 791
CC FT DOMAIN 54 59
CC FT DOMAIN 726 731
CC FT DOMAIN 742 747
CC FT DOMAIN 731 944
CC FT DOMAIN 944 AA; 111781 MW; 04FNA074BB8A0BC8 CRC64;
CC SQ SEQUENCE

Query Match 8.1%; Score 141; DB 1; Length 944;
Best Local Similarity 23.5%; Pred. No. 0.45;
Matches 61; Conservative 56; Mismatches 101; Indels 42; Gaps 8;

QY 129 MKQIQQFGENTRLHTAVENLK-----AVNVELSEQINLKQLTRLSDGDR 177
DB 129 MDLQKQKQESKRKLKDELNELETKFSENGSQSSAKENELKMLKNKIAELEEEISTKNSQ 375
QY 178 LEANTGDTALIAIDFQLSLEEFK-----SVGKVTMLSPFEKLAQSLK---ETFSQ- 226
DB 178 LIAEGKGLASLMA--QLTQLESKLNQRDSQLGSRSEELKKTNDKQDIRIAREEVSKD 433
QY 227 EAVQAMSSVTELRNLNALKELITENKTV-----IEQLKADAQLREEQVRFLE 275
DB 227 ERIDLQKKVQLENDLFVIKKTHSESKTITDNELESKDKLIKLENDLKVAGKYSKME 493
QY 276 KRKQLEPACSTLSHSIATQESTTLKSDTTLNHAVERSLIGVMVDGAESSTVEASQ 335
DB 276 K-----ELKKEREFNYKISESKLEDEKTLNKTISNLAENSQLNKI-----EDNSTATHMK 546
QY 336 DDSAQPDQENOSDAGEHKDS 355
DB 336 ENTEKQLESRLKDIETEKES 566
RESULT 3
EG5_HUMAN STANDARD; PRT; 1057 AA.
ID EG5_HUMAN
AC P52732; Q15716;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Kinesin-related motor protein Eg5 (Kinesin-like spindle protein HKSP)
DE (Thyroid receptor interacting protein 5) (TRIP5) (Kinesin-like protein
DE 1).
GN KNSLI OR EG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR Coiled coil; Phorbol-ester binding; SH3-binding.
FT NON_TER 1
FT DOMAIN <1 558
FT DOMAIN 365 561
FT DOMAIN 509 512
FT DOMAIN 522 670
FT DOMAIN 702 822
FT DOMAIN 851 1148
FT SITE 1212 1217
FT SEQUENCE 1286 AA; 146506 MW; 498101F9EA75E85 CRC64;
QY Query Match 8.3%; Score 143.5; DB 1; Length 1286;
Best Local Similarity 24.3%; Pred. No. 0.47;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
127 RHMKQ---IQQFGENTRLHTAVENLKAVNVELSEQI-----NOLKQLTRLRLS 172
DB 127 RNNKAQEEISELRQOKFYLETQAGKLEAQNKKLEEQLEKISHQDSDKNRLLELEFLRL 133
QY 173 DFGDRLAANTGDTALIAIDFQLSLEEFKSVGKVTMLSPFE-KLAQS---LKTF--SQ 226
DB 134 EVSLHEEQKLEKRLQTEQLSLQESQTLQALQAAALAESQLRQAKTELETTAEAE 193
QY 227 EAVQAMSSVTELRNLNALK-----LITENKTVIQLKAD----- 263
DB 194 EETQALTFAHDEIQRFKDALRNSCTVITDLEQLNQLTDNNAELNNQNFYLSKQLDEASG 253
QY 264 -----AQLREEQVRFLEKRRKQLEEEACSTLSHSIATFLOESTTLKSDTTLNHAVERSLI 317
DB 254 ANDEIVQLRSE-VDLHRLREITEREMOLTSOKOTMEALKTCTMLEEQVMDLEALNDELL 311
RESULT 2
NUF1_YEAST STANDARD; PRT; 944 AA.
ID NUF1_YEAST
AC P23380;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NUF1 protein (Spindle poly body spacer protein SPC110).
DE NUF1 OR SPC110 OR YDR356W OR D9476.3.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OS Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=92176232; PubMed=1541631;
RA Mirzayan C., Copeland C.S., Snyder M.;
RT "The NUF1 gene encodes an essential coiled-coil related protein that
RT is a potential component of the yeast nucleoskeleton."
RL J. Cell Biol. 116:1319-1332(1992).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=94064779; PubMed=7503995;
RA Kilmarlin J.V., Dyos S.L., Kershaw D., Finch J.T.;
RT "A spacer protein in the Saccharomyces cerevisiae spindle poly body
RT whose transcript is cell cycle-regulated."
RL J. Cell Biol. 123:1175-1184(1993).
RN [3]
RP SEQUENCE FROM N.A.
RP STRAIN=S288C; AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Ravello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Taich A., Trevaskis E., Vignati D.,

[illegible]

QY 121 TPOVVRHMKOO-----IQOFGENTRLHTA-VENLKAVN-----VELSEQINOLKQLHTR 170
 Db 1158 TSTQIELNKKREAEFLALRLDRLEATLQHEATVATLRKKHADSAELAEOQIDNLQRYK-- 1215
 QY 171 LSDFGDRLAANTGFTALIDFOLSLSEEFKSVGCTKVTMLSPFEKLAQSLKETFSQ----- 226
 Db 1216 -----QKLEKESEFKLEIDDLSSVE-----SVSKSKANLEKICRTLEDQLSEARGK 1263
 QY 227 -EAVQAMSSVVT-----ELRTNLNALKELITENKTVIOLKADAKALREEQVRFLEKQKQEL 281
 Db 1264 NEETQSLSELTTQKSLQTEAGELSKQLEEKESIVSQLSRSKQAFQTOIEELKQOLEE 1323
 QY 282 BEACSTLSHAIQLESTTLTKDSTNLHAVESRLIGVMVQDGAESSTVEEASQDSDAOP 341
 Db 1324 NKAKNALAHALQSSRHDCDLLREQYEEQBGKALQALSKANSEVAQWRTKYETDRIOR 1383
 QY 342 QDE 344
 Db 1384 TEE 1386

SULT 9 60_MOUSE

ID G160_MOUSE STANDARD; PRT: 1325 AA.
 AC P55937;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
 GN GOLGA3 OR MEA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Testis;
 RX MEDLINE=97217683; PubMed=9063644;
 RA Kondo M., Sutou S.;
 RT "Cloning and molecular characterization of cDNA encoding a mouse
 RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
 RL autoantigen.";
 RL DNA Seq. 7:71-82(1997).
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
 CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
 CC DETECTABLE MALE ANTIGEN (SDM).
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
 CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
 CC LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
 CC -!- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-19 OR MET-30 IS THE
 CC INITIATOR.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: D78270; BAA19612.1; -
 CC HSSP: P18852; ISCG.
 CC MGD; MGI:96958; Golga3.
 CC KW Spermatogenesis; Developmental protein.
 CC DOMAIN 201 204 POLY-ALA
 CC SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;

Query Match 7.8%; Score 136; DB 1; Length 1325;
 Best Local Similarity 23.6%; Pred. No. 1.3;
 Matches 67; Conservative 38; Mismatches 101; Indels 78; Gaps 9;

QY 126 VRHMKOIQO-----FGEENTRL-----HTAV-----ENLKAVNVLSQIN----- 162
 Db 405 LQHARWYQQOQLAQEARVRLOGEAMHIQVQOMTQAGLLEHLKLENVSLSHQLTETQHR 464
 QY 163 -----QKQLHTRLSD-----FGDRLAANTGFTALIDFOLSLSEEFKSVGCTKVT 208
 Db 465 SIKERITAVQLQSLQSTEADMLDQEAFAVQIREAK-----TWVEDLQRRLEEFEGEREQLOK 520
 QY 209 MLSPEKLAOSLKE-----TFQBAVQAMSSVTELRN 242
 Db 521 VADAASLEQQLQEVKVLTLFORDQALAAQOEHLVDVIQLTSTOBALQAKGSLDLDLHTR 580
 QY 243 LNALKELITENKTVIOLKADAKALREEQVRFLEKQKQLEEACSTLSHAIQLESTTL 302
 Db 581 VDEL-----QARLEELQREADSREDAIHFLQNEKIVLEVALQSAKSDKEELDORGARRL 633
 QY 303 KDSNTNLHAVESRLIGVMVQDGAESSTVEEASQDSDAOPDENQ 346
 Db 634 EEDTEE-----TSGLLQLRODLAVKSNQVHEHLOQETATLRKQM 673

RESULT 10 MYH8_HUMAN

ID MYH8_HUMAN STANDARD; PRT: 1937 AA.
 AC P13535; Q14910;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).
 GN MYH8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90323631; PubMed=2373371;
 RA Karsch-Mizrachi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;
 RT "Generation of a full-length human perinatal myosin heavy-chain-
 RT encoding cDNA.";
 RL Gene 89:289-294(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=95324556; PubMed=7601129;
 RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
 RA Stedman H.H., Rubinstein N.A.;
 RT "Characterization of a human perinatal myosin heavy-chain
 RT transcript.";
 RL Eur. J. Biochem. 230:1001-1006(1995).
 RN [3]
 RP SEQUENCE OF 502-1937 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90235862; PubMed=1691980;
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [4]
 RP SEQUENCE OF 860-1937 FROM N.A.
 RX MEDLINE=89234168; PubMed=2715179;
 RA Feghali R., Leinwand L.A.;
 RT "Molecular genetic characterization of a developmentally regulated
 RT human perinatal myosin heavy chain.";
 RL J. Cell Biol. 108:1791-1797(1989).
 RN [5]
 RP SEQUENCE OF 1-46 FROM N.A.
 RA Esser K., Tidhar A., Myszkowski M.;
 RT "Isolation and characterization of the human perinatal MHC promoter.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MUSCLE CONTRACTION.

QY 130 KOIOOFGE---ENTRLHTAVENKAVNVELSEIQNOLKOLHTRLSDFGDRLEA----- 180
 DB 1076 KOOLDEKLEKKEFEISNLISKIEDQAVEIQLOKKI---KELQARIEELGEEIEAERASR 1132
 QY 181 -----NTGFTALADFOLSEEFKSVGKVTML-----SPFEKLAOSLKE-TFSQAV 229
 DB 1133 AKAEQRSDLSRELEIEISERLEE-AGGATSAQVELNKKREAEQFKLRDLLEATLOHEAM 1191
 QY 230 QA-----MMSSVTELTNTNALKELITENKTIEQKADAKAQLREQVRFLEKRKQEL 281
 DB 1192 VAALRRKHADSAEAGEQIDNLRVQKLEKEKS---ELKMETDSSLSSNAEAIKAGNUL 1248
 QY 282 EEACSTLSHSIATL-----QESTTLKSDTNTNHVESRLIGVMVQDGAESSTVEASQDD 337
 DB 1249 EKMCRSLEDOVSELKTEEBEQRLINDLTAQARLQTEAGEYSRQLEDKDALVLSQLSRSK 1308
 QY 338 SAQPQ 342
 DB 1309 QASTQ 1313

RESULT 11

USOL_YEAST STANDARD; PRT; 1790 AA.
 ID USOL_YEAST
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 GN USOL OR INT1 OR YDL058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-X2180-1A;
 RX MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yanasaki M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae."
 RT J. Cell Biol. 113:245-260(1991).
 RL [2]
 RN SEQUENCE OF 782-1790 FROM N.A.
 RP Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M36769; AAC17185.1; -;
 CC EMBL; Z38133; CAA86293.1; -;
 CC EMBL; X51592; CAA35941.1; -;
 CC EMBL; M35250; AAA36346.1; -;
 CC EMBL; AF067143; AAC21557.1; -;
 CC PIR; A30220; A30220.
 CC HSSP; P08799; 1LVK.
 CC MIM; 160741; -;
 CC InterPro; IPR000048; IQ.
 CC InterPro; IPR004009; Myosin_N.
 CC InterPro; IPR002928; Myosin_tail.
 CC InterPro; IPR001609; myosin_head.
 CC Pfam; PF00612; IQ; 1.
 CC Pfam; PF00063; myosin_head; 1.
 CC Pfam; PF02736; Myosin_N; 1.
 CC Pfam; PF01576; Myosin_tail; 1.
 CC PRINTS; P00193; MYOSINHEAVY.
 CC ProDom; PD000355; myosin_head; 1.
 CC SMART; SM00015; IQ; 1.
 CC SMART; SM00242; MYSC; 1.
 CC PROSITE; PS50096; IQ; 1.
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 Calmodulin-binding. 780 MYOSIN HEAD-LIKE.
 DOMAIN 1
 DOMAIN 781 813
 DOMAIN 842 1937
 NP_BIND 181 188
 DOMAIN 658 680
 DOMAIN 760 774
 DOMAIN 132 132
 MOD_RES 698 698
 MOD_RES 708 708
 CONFLICT 15 15
 CONFLICT 970 970
 CONFLICT 1072 1072
 CONFLICT 1247 1247
 CONFLICT 1251 1252
 CONFLICT 1261 1261
 CONFLICT 1297 1297
 CONFLICT 1377 1378
 CONFLICT 1504 1505
 CONFLICT 1847 1847
 CONFLICT 1914 1914
 SEQUENCE 1937 AA; 222762 MW; A3EE2D151792E9E8 CRC64;
 Query Match 7.8%; Score 135.5; DB 1; Length 1937;
 Best Local Similarity 22.9%; Pred. No. 2.2;

-!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC-2)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 SHOWING
 -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 -!- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M36769; AAC17185.1; -;
 CC EMBL; Z38133; CAA86293.1; -;
 CC EMBL; X51592; CAA35941.1; -;
 CC EMBL; M35250; AAA36346.1; -;
 CC EMBL; AF067143; AAC21557.1; -;
 CC PIR; A30220; A30220.
 CC HSSP; P08799; 1LVK.
 CC MIM; 160741; -;
 CC InterPro; IPR000048; IQ.
 CC InterPro; IPR004009; Myosin_N.
 CC InterPro; IPR002928; Myosin_tail.
 CC InterPro; IPR001609; myosin_head.
 CC Pfam; PF00612; IQ; 1.
 CC Pfam; PF00063; myosin_head; 1.
 CC Pfam; PF02736; Myosin_N; 1.
 CC Pfam; PF01576; Myosin_tail; 1.
 CC PRINTS; P00193; MYOSINHEAVY.
 CC ProDom; PD000355; myosin_head; 1.
 CC SMART; SM00015; IQ; 1.
 CC SMART; SM00242; MYSC; 1.
 CC PROSITE; PS50096; IQ; 1.
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 Calmodulin-binding. 780 MYOSIN HEAD-LIKE.
 DOMAIN 1
 DOMAIN 781 813
 DOMAIN 842 1937
 NP_BIND 181 188
 DOMAIN 658 680
 DOMAIN 760 774
 DOMAIN 132 132
 MOD_RES 698 698
 MOD_RES 708 708
 CONFLICT 15 15
 CONFLICT 970 970
 CONFLICT 1072 1072
 CONFLICT 1247 1247
 CONFLICT 1251 1252
 CONFLICT 1261 1261
 CONFLICT 1297 1297
 CONFLICT 1377 1378
 CONFLICT 1504 1505
 CONFLICT 1847 1847
 CONFLICT 1914 1914
 SEQUENCE 1937 AA; 222762 MW; A3EE2D151792E9E8 CRC64;
 Query Match 7.8%; Score 135.5; DB 1; Length 1937;
 Best Local Similarity 22.9%; Pred. No. 2.2;

DR EMBL; L03188; AAB00143.1; -;
 DR EMBL; U53668; AAB66659.1; -;
 DR PIR; A38455; A38455;
 DR HSSP; P80220; 1DIP.
 DR SGD; S0002216; USOI.
 DR InterPro; IPR002017; Spectrin.
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 FT SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
 Query Match 7.7% Score 134; DB 1; Length 1790;
 Best Local Similarity 21.3%; Pred. No. 2.5;
 Matches 57; Conservative 53; Mismatches 121; Indels 36; Gaps 8;
 QY 117 IRDTPSOVVRHMKQIQOFGENTRLHTAVENLKAVNVE-----LSEQ 160
 Db 1372 LNEGS-STTQYSEKINTLEDELIRLQNEE-LKAKEIDNTRSELEKVSLSNDELLEEK 1429
 QY 161 INQLQLHLTRLSDFGRLEANTGDTALADFDLSLEEFK-----SVGTVETMLSPF 213
 Db 1430 QNTIKSLQELSYKDKITRNDKLSIERDNKRDLESKEQLRAAQESKAKVEEGLKL 1489
 QY 214 EKLAQSLKETF----SOEAYQAMKSSVTELTNLNALKELITENKTVIQLKADAOLREE 269
 Db 1490 EE--ESSKEKAELEKSKEMMKLESIESNETELKSMETIRKSDKLEQSKSA---EE 1544
 QY 270 QVRFLEKRLKELEACSTLSHSTATQESTTLTKDSTNLHAVESRLIGV--VODGAS 327
 Db 1545 DINLQHEKSDLSIRNESEKDEELKSLRIEAKSGSELTVKQLNNAQEKIRINAE 1604
 QY 328 STVEEASQDASQPDENSDAGEKD 354
 Db 1605 NTVLKSLEIDIERELKDKQAEIKSNOE 1631

RESULT 12

MYH4_HUMAN
 AC MYH4_HUMAN STANDARD; PRT; 1939 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, fetal (Myosin heavy chain IIB)
 DE (MyHC-IIB).
 GN MYH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Skeletal muscle;
 RX MEDLINE=9318869; PubMed=10388558;
 RA Wells A., Schiaffino S., Leinwand L.A.;
 RT "Comparative sequence analysis of the complete human sarcomeric myosin
 heavy chain family: implications for functional diversity";
 RL J. Mol. Biol. 290:61-75(1999).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).
 EMBL; AF111783; RAD29949.1; -;
 DR HSSP; P08799; IMMD.
 DR MIM; 160742; -;
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS0096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 814 IQ
 FT DOMAIN 843 1939 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT SEQUENCE 1939 AA; 223012 MW; 40BIAD1D77A47DE CRC64;

Query Match 7.7% Score 133.5; DB 1; Length 1939;
 Best Local Similarity 20.4%; Pred. No. 2.9; Indels 25; Gaps 5;
 Matches 52; Conservative 55; Mismatches 55; Gaps 5;
 QY 121 TPQOVVRHMK-----QIQOFGENTRLH----TAVENLKAVNVE-ELSEQINLKOLHTR 170
 Db 1161 TSAQIELNKKREAFQKMRDLEESTLQHEATAAALRKKHADSVAEELGKQIDSLQVKK 1220
 QY 171 LSDFGDRLEANTGDTALADFDLSLEEFKSVGTVETMLSPFEKRLKELEACSTLSH 230
 Db 1221 LEKEKSELKWEINDLASNMETVSKANFEKMCRTLEDQLSEIK-----TKEEQORLIN 1275
 QY 231 AMSSVTELTNLNALKELITENKTVIQLKADAQALREQVRFLEKRLKELEACSTLSH 290
 Db 1276 ELQAQKARLHTESEGFQRDLDEKDMVLSGRKQAFQIIEELKQLKEETKAKSTLAH 1335
 QY 291 STATQESTTLTKDSTNLHAVESRLIGVVDGAEESTVEASQDSDSAQPDENQ----346
 Db 1336 ALQSAHDCDLLREQVEEQEAKAEALQKMGSKANSEVAQWRKYETDAIQTEELEAKK 1395
 QY 347 -----SDAGEHKDS 355
 Db 1396 KLAQRLQDAEEHVEA 1410

Search completed: August 13, 2002, 09:30:42
Job time: 378 sec

Result No.	Score	Query Match	§			Description
			Length	DB	ID	
1	140	8.1	1057	4	US-09-541-782-10	Sequence 10, Appl
2	131	7.6	2482	1	US-08-328-254-6	Sequence 6, Appl
3	131	7.6	3248	1	US-08-353-700-1	Sequence 1, Appl
4	131	7.6	3248	5	PCR-US95-16216-1	Sequence 1, Appl
5	126	7.3	2101	1	US-08-466-390-4	Sequence 4, Appl
6	126	7.3	2101	1	US-08-470-950-4	Sequence 4, Appl
7	126	7.3	2101	1	US-08-467-781-4	Sequence 4, Appl
8	126	7.3	2101	1	US-08-195-487-4	Sequence 4, Appl
9	126	7.3	2101	2	US-08-483-924-4	Sequence 4, Appl
10	126	7.3	2101	4	US-09-452-294-1	Sequence 4, Appl
11	126	7.3	2101	5	PCR-US93-06160-4	Sequence 1, Appl
12	125.5	7.2	1184	4	US-09-541-782-2	Sequence 4, Appl
13	122.5	7.1	580	2	US-08-591-079-2	Sequence 2, Appl
14	119	6.9	976	4	US-09-104-324B-4	Sequence 2, Appl
15	118	6.8	180	1	US-08-328-254-7	Sequence 2, Appl
16	117.5	6.8	1388	2	US-08-685-576-1	Sequence 4, Appl
17	117	6.8	1312	2	US-08-592-126-148	Sequence 7, Appl
18	117	6.8	1312	2	US-08-687-080-51	Sequence 1, Appl
19	116	6.7	576	2	US-08-533-306A-2	Sequence 148, App
20	116	6.7	576	2	US-08-742-923A-2	Sequence 51, Appl
21	113.5	6.5	1068	4	US-09-085-199B-11	Sequence 2, Appl
22	113.5	6.5	1354	3	US-08-685-871-2	Sequence 2, Appl
23	112	6.5	546	2	US-09-067-351-1	Sequence 11, Appl
24	112	6.5	546	4	US-09-360-490-1	Sequence 1, Appl
25	112	6.5	1388	2	US-08-685-576-4	Sequence 4, Appl
26	109	6.3	896	1	US-08-085-737-2	Sequence 2, Appl
27	109	6.3	896	1	US-08-480-145-2	Sequence 2, Appl

TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
; PS-08-353-700-1

```

Query Match          7.6%; Score 131; DB 1; Length 3248;
Best Local Similarity 21.9%; Pred. No. 0.016;
Matches 64; Conservative 44; Mismatches 92; Indels 92; Gaps 10;

140 NTRLHTAVENLKA VNVLSQINQLKOLH-----TRISDFGRLEANTGDTTALIADPQLS 195
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

846 NSIDLQOCELVQIKGEIENLMAEHOSFVAETSQRISKIQEDTSAHONVVAETLSA 905
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

196 LE-----EFKSVCYKVEHMLSPFEKLAQS-----LKEFFSQEAQVAAQMSVVTE 238
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

906 LENKELOLNDKVEETEQAIEQLKKSNNHLLDSLKELOLSETLSLEKKE--MSSIIS 963
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

239 LRTNLALKELITENKTVIEQLKADALREQVRFLEK-----276
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

964 L--NKREIETLOENGTLKE--INASLNQEKMLLIQKSFANYIDEREKSIELSDQY 1018
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

277 -----BQOLEEACSTLSHSIATLQESTT-----LLKDS-----306
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1019 QKEKLILLQCEETGNAYEDLSQYKAAQEKNSKLECLLNCTSLCENKRNKEQLKEAF 1078
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

307 -----TNLHAVESRLIGVWODGAESSTVEEASODDSAQPDENQSDAG 350
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1079 AKEHOEFLTKLAAFEERNQNLML-----ELETVQQAALREMTDNNQNSKSEAG 1126
      QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      4
PCT-US95-16216-1
; Sequence 1, Application pc/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4

```

GENERAL INFORMATION:
 APPLICANT: Zhu, Xueliang
 APPLICANT: Lee, Wen-hwa
 TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328,254
 FILING DATE: 24-OCT-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/141,239
 FILING DATE: 22-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 1191
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2482 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-328-254-6

```

Query Match      7.6%; Score 131; DB 1; Length 2482;
Best Local Similarity 21.94%; Pred. No. 0.011; 92; Indels 92; Gaps 10;
Matches 64; Conservative 44; Mismatches 9;

140 NTRLHTAVENLKAVNVELSQINQLKOLH---TRUSDGDRLEANTGDTTALIADPOL 195
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
214 NSDLQKQCEELVQIKGEIENLMAEQMHQSFAVETSRISKLOEDTSAHQVWVAETLSA 273
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
196 LE----EPKSVGTGVETMLSPFEKLAQS-----LKEFTSQEAVQAMVSVYTE 238
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
274 LENKEBELLOLLNDKVETEQAQIEQLKKSNHLLSDSKELQLLSETLSLEKKE--ASSIIS 331
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
239 LRTNLNALKLITENTKTVIEOLKADADQLREOVRFLEK----- 276
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
332 L--NKREIEELTQENGTLKE---INASLNQEKWNLQKSESFANYIDREKSISELSDQY 386
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
277 -----RKQLEEARCSTLSHSIATLQSTT-----LLKDSF----- 306
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
387 KOEKUILLQRCETGNATVEDYQYKAAQEKNSKLECLLNECTSLCENRKNELEQLKEAF 446
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
307 -----TNLHAVESRLIGYVQDGAESPSTVEEASODDSAQPDQBNQSDAG 350
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
447 AKPHOPFTTKTAFAPERNOMLML-----ELETVOALRSEMFTDNNKSKSEAG 494
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

1  RESULT 3
2  US-08-353-700-1
3  : Sequence 1, Application US/08353700
4  : Patent No. 559919
5  : GENERAL INFORMATION:
6  : APPLICANT: YEN, TIMOTHY J.
7  : APPLICANT: RATTNER, JEROME B.
8  : TITLE OF INVENTION: NUCLEIC ACID ENCODING A
9

```


COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4

Query Match 7.3%; Score 126; DB 1; Length 2101;
Best Local Similarity 22.5%; Pred. No. 0.024;
Matches 55; Conservative 48; Mismatches 97; Indels 44; Gaps 9;
QY 128 HMKQIQOFGENTRLHTAVENLKAVNVELSEQINQLKQLHTRLSDFGDRLEAN-----TG 183
DB 456 HFEKQQLSSLITDQSSISNLS-----QAKEELEQASQAH-----GARLTAQVASLTS 505
QY 184 DFTALIADQFLSLEEFKSVGTVETMLSPFELKLAQSLKETFQAEVQAMSSVTELRNL 243
DB 506 ELTTNATIOQDOELAGLKKQAKQKAQ-----QLAQTLOQ--QEQASQGLRHQVEQLS--- 557
QY 244 NALKELITENKTVIEQLKADQLRREQVRFLKRRKQLEEFACSTLSHSIATLQESTTLK 303
DB 558 SSLKQEQQLKEVAEKQEAETRDHAQOL-----ATAAEEREASLRERDAALKOLEALEK 611
QY 304 DSTTNLHVESRL-IGVMVQDGAESSTVE-----EASODDSAQPDENOS 347
DB 612 EKAACLEILQOQLQVANEARDSAQTSTVQAQREKAEKSRKVEELQACVETARQEQHEAQA 671
QY 348 DAGE 351
672 QVAE 675

RESULT 7
US-08-467-781-4
Sequence 4, Application US/08467781
Patent No. 5780596
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467.781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-781-4

Query Match 7.3%; Score 126; DB 1; Length 2101;
Best Local Similarity 22.5%; Pred. No. 0.024;
Matches 55; Conservative 48; Mismatches 97; Indels 44; Gaps 9;
QY 128 HMKQIQOFGENTRLHTAVENLKAVNVELSEQINQLKQLHTRLSDFGDRLEAN-----TG 183
DB 456 HFEKQQLSSLITDQSSISNLS-----QAKEELEQASQAH-----GARLTAQVASLTS 505
QY 184 DFTALIADQFLSLEEFKSVGTVETMLSPFELKLAQSLKETFQAEVQAMSSVTELRNL 243
DB 506 ELTTNATIOQDOELAGLKKQAKQKAQ-----QLAQTLOQ--QEQASQGLRHQVEQLS--- 557
QY 244 NALKELITENKTVIEQLKADQLRREQVRFLKRRKQLEEFACSTLSHSIATLQESTTLK 303
DB 558 SSLKQEQQLKEVAEKQEAETRDHAQOL-----ATAAEEREASLRERDAALKOLEALEK 611
QY 304 DSTTNLHVESRL-IGVMVQDGAESSTVE-----EASODDSAQPDENOS 347
DB 612 EKAACLEILQOQLQVANEARDSAQTSTVQAQREKAEKSRKVEELQACVETARQEQHEAQA 671
QY 348 DAGE 351
672 QVAE 675

RESULT 8
US-08-195-487-4
Sequence 4, Application US/08195487
Patent No. 5783403
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 08/195,487
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,701
FILING DATE:

```

: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ. EDMUND R
: REGISTRATION NUMBER: 27, 829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7000
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-195-487-4

```

Query Match	7.3%;	Score 126;	DB 1;	Length 2101;
Best Local Similarity	22.5%;	Pred. No. 0,234;		
Matches 55;	Conservative 48;	Mismatches 97;	Indels 44;	Gaps
128	HKQOIQOQGEENTRLHTAVENLKVAVNELSEQINQLKQLHTRLSDFGDRLLEAN----	TG 183		
456	HFEEEKQQLSLITDQSSISNLS-----QAKEEEQASQAH-----	GARLTAQAVASLTS 505		
184	DFTALIHADFOLSEEPKSVGTVETMLSPFFKFLQAQSLKETFQBAVQAMMSSVTELRTNL 243			
506	ELATFLNATIQOQDQELAGLQKAQEKQA-----QLAQTILQQ--QEQASGLRHQVEQLS----	557		
244	NALKELITENTKVIEQLKADAOALREEQVRFLEKKRQBLEACSTLSHTATIQESTTLK 303			
558	SSLKREQEQLKVEAKQEAQETRDHQAQL-----ATAAEEREASLERDAALKQLEALPK 611			
304	DSTNTNLHVESRL-IGVMVQDQAGBSSTVE-----EASQDSDAQPODENQS 347			
612	EKAAKLEILQOOLQVANEARDSAQTTSVTQAQREKAELSRKVEELQACVETARQEQHEQA 671			
348	DAGE 351			
672	QVAE 675			

RESULT 9
 US-08-483-924-4
 ; Sequence 4, Application US/08483924
 ; Patent No. 5882876
 ; GENERAL INFORMATION:
 ; APPLICANT: TOUTKATLY, GARY
 ; APPLICANT: LIGDARD, GRAHAM P
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: 125 HIGH STREET
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/483,924
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PITCHER ESQ, EDMUND R
 ; REGISTRATION NUMBER: 27,829
 ; REFERENCE/DOCKET NUMBER: MTP-013
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 248-7000

```

; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-924-4

```

Query Match	7.3%	Score 126;	DB 2;	Length 2101;
Best Local Similarity	22.5%	Pred. No. 0.024;		
Matches	55;	Conservative 48;	Mismatches 97;	Indels 44; Gaps 9;
128	HKMQIQOQFGBENTRLHTAVENLKAVNVELSEQINOLKOLHRLSDPFGDRLEAN----	TG 183		
456	HFEEKQQLSLITDQLQSSINLS-----QAKKELEQASQAH-----	CARLTAQVASTS 505		
184	DFTALIADFQLSEEFKSVGKVTETMLSPFEKLAOSLKTFFSQEAVQAMMSVTELRNL 243			
506	ELTTLNATIQQQOQELAGLKQAKAQQA---QLAQTQQ--QEQASQGLRHQVEQLS---	557		
244	NALKELITENKTVIEQLKADAQRLREQVFELEKRRKQEEACSTLHSHATLQESTILLK 303			
558	SSLKQEQQLKVAEYKQEQATRDQHAQQL-----ATAEEREASLRERDAALKQLEALEK 611			
304	DSITNTHAVESRL-IGWVQDQAGNESTVE-----EASODDSAQPDENQS 347			
612	EKAUKETLQQQLQVANEAREDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQA 671			
348	DAGE 351			
672	QVAE 675			

```

RESULT 10
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lellievre, Sophie
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPH AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1

```

```

Query Match      7.3%; Score 126; DB 4; Length 2101;
Best Local Similarity 22.5%; Pred. No. 0.024;
Matches 55; Conservative 48; Mismatches 97; Indels 44; Gaps 9;

QY   128  HKQQIQOQGGEENTRLHTAVENLKAVNVLSQINQLKHLTRLSPGDRLEAN---TG 183
      |::|||:::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    456  HFEEKQQQLSLTDTQSSISNL-----QAKEBLEAQSAH-----GARLTAAVASLTS 505
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   184  DETALLADPQLSEDEEFSKVGTVETWLSPFPEKLQASIKETFEOEAQVAMMSSYTELRTNL 243
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    506  ELTTLNATIQQDOOLAGLKKQAKEKA---QLAQTQQ--SQASOGILRHQVEQLS--- 557
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   244  NALKELITENVIVQELKADAQALREQRVRFLKRKQOLEBEACSTLSHSIATLOESTTLK 303

```

Tue Aug 13 14:23:42 2002

Db 558 SLLKQEQQLKEVAEQEATROHAQOL-----ATAAEEREASLRERDAALKQLEALEK 611
QY 304 DSTTNLHAVESRL-IGVMVDGAESESTVE-----EASODDSAQPODENOS 347
Db 612 EKAACLEILQOOLQVANEARDSAQTSVTOAQREKAEKSRKVBELQACVETARQEQHEAQA 671
QY 348 DAGE 351
Db 672 QVAE 675
RESULT 11
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-06160-4
Query Match 7.3%; Score 126; DB 5; Length 2101;
Best Local Similarity 22.5%; Pred. No. 0.024;
Matches 55; Conservative 48; Mismatches 97; Indels 44; Gaps 9;
QY 128 HMKQOIQQFGEENTRLHTAVENLKAVNVLSLQINLKQLHTRLSDFGDRLEAN-----TG 183
Db 456 HFEEKQQLSSLLTDLQSSISNLS-----QAKBELEQASQAH-----GARLTAQVASLTS 505
QY 184 DFTALADFOLSLEEFKSVGTVMSPFKLAQSLKFTFQEAQVQAMMSVSTLELRTNL 243
Db 506 ELATLNATIQOQDELQAGLQKQAKQQA-----QLAQTLLQ--QEQAQGLRHQVEQLS--- 557
QY 244 NALKELITENKTVIOLKADAQLRERQVRFLKRRKQLEFEACSTLSHSIATLQESTTLK 303
Db 558 SLLKQEQQLKEVAEQEATROHAQOL-----ATAAEEREASLRERDAALKQLEALEK 611
QY 304 DSTTNLHAVESRL-IGVMVDGAESESTVE-----EASODDSAQPODENOS 347
Db 612 EKAACLEILQOOLQVANEARDSAQTSVTOAQREKAEKSRKVBELQACVETARQEQHEAQA 671
QY 348 DAGE 351

Db 672 QVAE 675
RESULT 12
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: A. nidulans
; US-09-541-782-2
Query Match 7.2%; Score 125.5; DB 4; Length 1184;
Best Local Similarity 19.8%; Pred. No. 0.011;
Matches 49; Conservative 59; Mismatches 111; Indels 29; Gaps 7;
QY 119 DGTFSQVVRHMKQIQFGEENTRLHTAVENL-----KAVNVLSLQINLKQLHTR 170
Db 697 NGLSAAARLSEEVIGFTQLHSQLSFTNNLGLKDLKLSIFETMATHLSQKNEINLRAE 756
QY 171 LSDFGDRLEANTGFTALADFOLSLEEFKSVGTVMSPFKLAQSLKFTFQEAQV 230
Db 757 LQSSNRQNIETTHKASHLA--QAIEEHVAAERREILMSQIKALVEESR----QKQFA 810
QY 231 AMMSVTELTNLNALKELITENKTVIOLKADAQLR-----EQVRFLKRRKQLEFEACST 287
Db 811 RLRAKIDGVVTEISASGDMLEQATT-----QHDRQIDEWVFKSEQFAKDVNASKDEIRTK 865
QY 288 LSHSIATLQESTTLKDKSTNLHAVESRLIGVMVDGAESESTVEEASODDSAQPODENOS 347
Db 866 LQNDWEAFDQORNSTIRKATESVHKETVRIYDVQVDDMGROM---EALDDFVARARSON-- 920
QY 348 DAGEKDS 355
Db 921 --GRYRDA 926
RESULT 13
US-08-591-079-2
; Sequence 2, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinksky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,079

;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Livnat, Shmuel
;; REGISTRATION NUMBER: 33,949
;; REFERENCE/DOCKET NUMBER: 15661-20017.00
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 887-1500
;; TELEFAX: (202) 887-0764
;; TELEX: 90-4030 MRSNFOERSH
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 580 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-591-079-2

Query Match
Best Local Similarity 7.1%; Score 122.5; DB 2; Length 580;
Matches 54; Conservative 51; Mismatches 98; Indels 35; Gaps 9;
QY 123 SOWRHMKQIOQFGEENTRLHTAVENLKAVNVSEIQINOLKQLHTRLSDFGDRLEANT 182
DB 77 SUTLLIGNLIQILGKSL---TALTN-KITAWKSQQQAQQKNL-----EFSKINTLL 127
QY 183 GDTALIAADPQLSLEEFKSVGTQKVTMLSPFEKLAQSL-----KETFSQAVQMM 233
DB 128 SETGLTRDYEQINKLNADSKIDLENKINQTRLSNLDPSPEKKLSREIQ--- 184
QY 234 SSVTELRNLNALKELITENTVIEQ---LKADAQLREBQVRFLEKRRQLEAEACSTLS-H 290
DB 185 -----LTIKDAAVKDTLIEQKTLSTHSKLTOKSMQ-LEKEIDTSFAPSNTASAE 234
QY 291 SIATLQESTLLKDDSTNLHVESRLGVMVQDGAESTVEEASQDSDSAPQDENQSD 348
DB 235 QLSQKSLTGLA-SVTOLMATFQLVGVKNNEESLKNLDLALFQSLQESRKTEMERKSD 291

RESULT 14
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T. Recl, Ozlem, Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: Wordperfect
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104.324B
; FILING DATE: 25-June-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/892,702
; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (212) 318-3000
;; TELEFAX: (212) 752-5958
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 976 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
US-09-104-324B-4

Query Match
Best Local Similarity 6.9%; Score 119; DB 4; Length 976;
Matches 60; Conservative 37; Mismatches 80; Indels 78; Gaps 10;
QY 139 ENTRLHTAVENLKAVNVSEIQINOLKQLHTRLSDFGDRLEANTGDTALIAADPQLSLEE 198
DB 394 EQORLEKNEQDKILTWELOKKSSELEEM-----TKLTNNKEVELEE 435
QY 199 FKSIVGTQKVTML---SPFEKLAQSLKETFSQEA---QAMSSSVTELRNLNAL----- 246
DB 436 LKKVGEKETLLYENKQFEXIAEELKGT-EQELIGLLOAREKEVHDLLEIOLTAITTSQY 494
QY 247 -----KELITE-----NKTVIE-----OLKAD--AQLREBQVRFLEKRRQ 279
DB 495 YSKEVKDKLQTELENEKLNKTELTAHSHCNKLSLENKELTQETSDMTLELNQOEDINNNKKQ 554
QY 280 ELEBACSTLSHSTATLQESTTLLKDS-----TTNLHVESRLGVMVQDQ 324
DB 555 E-----ERMLKQLENQETQETQLRNELEYVREELKQKRDVEKCKLDKSENCNLRKQVE 609
QY 325 AESTVEEASQDSDA 339
DB 610 NKNKYIEELQOENKA 624

RESULT 15
US-08-328-254-7
; Sequence 7, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 180 amino acids

Tue Aug 13 14:23:42 2002

```

; TYPE: amino acid
; TOPOLOGY: linear
US-08-328-254-7

```

Query Match	6.8%	Score 118;	DB 1;	Length 180;
Best Local Similarity	24.0%	Prod. No. 0.0033;		
Matches	41;	Conservative	47;	Indels 42; Gaps 7;
QY	128	HKQOIQIQGENTPLTAVENKAVNVELSEGINOLKQLHTRLSDFGDRLEANTGDFTA	187	
		: : : : : : : : : : :		
Db	26	HIAEKLKERENOSLKDKNLER-ELQMSENQEL-----	61	
		: : : : : : : : : : :		
QY	188	LIADFQLSLREFFKSVGTVETMLSPFFKLQAOSLKTFTSOEAV-----QAMKSSVTEI	241	
		: : : : : : : : : : : : : :		
Db	62	IVLDAENSKAE-----VETLKTQIEEMARSLK-VFELDLVTLRSEKENLTKQIQEKQ	113	
		: : : : : : : : : : : : : :		
QY	242	NLNAUKELITENKTVIEQL-KADQALRREQ---VRFLERKQKELEACSTL	288	
		: : : : : : : : : : : : : :		
	114	QISELDKLLSSFLKSEKEQAQIQKEESTAVEMQLNQKLKEINVAAL	164	
		: : : : : : : : : : : : : :		

Search completed: August 13, 2002, 09:24:18
Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:24:04 ; Search time 48.57 Seconds
(without alignments)
1264.427 Million cell updates/sec

Title: US-09-673-763-8

Perfect score: 1733

Sequence: 1 MTVSTNDTSPVISRASSPTF.....DSSAQPDENQSDAGEHKOS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	2 Q46210	Q46210 chlamydia t
2	448	25.9	225	2 Q53263	Q53263 chlamydia p
3	200	11.5	390	16 Q92828	Q92828 chlamydia p
4	185	10.7	276	16 Q9PKR8	Q9PKR8 chlamydia m
5	166.5	9.6	273	2 Q9RFK7	Q9RFK7 chlamydia t
6	164.5	9.5	273	2 Q9AMA9	Q9AMA9 chlamydia t
7	164.5	9.5	273	2 Q9AM94	Q9AM94 chlamydia t
8	164.5	9.5	273	16 Q84121	Q84121 chlamydia t
9	163.5	9.4	273	2 Q99Q56	Q99Q56 chlamydia t
10	162.5	9.4	273	2 Q9F7K9	Q9F7K9 chlamydia t
11	162.5	9.4	273	2 Q69196	Q69196 chlamydia t
12	162.5	9.4	273	2 Q9AMA7	Q9AMA7 chlamydia t
13	161.5	9.3	273	2 Q9AMB2	Q9AMB2 chlamydia t
14	161.5	9.3	273	2 Q9AMA8	Q9AMA8 chlamydia t
15	161.5	9.3	273	2 Q9AMA6	Q9AMA6 chlamydia t
16	160.5	9.3	273	2 Q9AMB1	Q9AMB1 chlamydia t

17	160.5	9.3	273	2 Q9AMA4	Q9AMA4 chlamydia t
18	157.5	9.1	273	2 Q9AMB0	Q9AMB0 chlamydia t
19	157.5	9.1	273	2 Q9AMA5	Q9AMA5 chlamydia t
20	150.5	8.7	1627	5 Q96200	Q96200 giardia lam
21	149	8.6	256	2 Q9F7L1	Q9F7L1 chlamydia t
22	149	8.6	533	17 Q9YCP2	Q9YCP2 aeropyrum p
23	149	8.6	612	4 Q96JV2	Q96JV2 homo sapien
24	148.5	8.6	568	10 Q932J6	Q932J6 arabidopsis
25	146.5	8.5	924	5 Q15738	Q15738 dictyosteli
26	145.5	8.4	1732	11 Q54874	Q54874 rattus norv
27	145	8.4	409	16 Q9CKU8	Q9CKU8 pasteurella
28	145	8.4	1690	5 Q4929	Q4929 drosophila
29	145	8.4	1690	5 Q9VJE5	Q9VJE5 drosophila
30	143.5	8.3	1618	11 Q9QX19	Q9QX19 rattus norv
31	143	8.3	1410	4 Q14221	Q14221 homo sapien
32	143	8.3	1411	4 Q15075	Q15075 homo sapien
33	143	8.3	1935	5 Q4934	Q4934 loligo peal
34	141	8.1	678	2 Q9KK49	Q9KK49 streptococ
35	140.5	8.1	403	16 Q9X252	Q9X252 thermotoga
36	140.5	8.1	593	10 Q9AS76	Q9AS76 oryza sativ
37	140	8.1	446	4 Q9RCH4	Q9RCH4 homo sapien
38	140	8.1	770	16 Q9KKP6	Q9KKP6 vibrio chol
39	139.5	8.0	739	17 Q9UYF8	Q9UYF8 pyrococcus
40	139.5	8.0	746	5 Q25561	Q25561 naegleria f
41	138.5	8.0	1641	11 Q88528	Q88528 mus musculu
42	138.5	8.0	1762	10 Q94DC2	Q94DC2 oryza sativ
43	138.5	8.0	1957	5 Q04010	Q04010 onchocerca
44	138.5	8.0	2055	11 Q88938	Q88938 mus musculu
45	137	7.9	592	2 Q00720	Q00720 group g str

ALIGNMENTS

RESULT 1
ID Q46210 PRELIMINARY; PRT: 355 AA.
AC Q46210;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE LOCALISED PROTEIN.
GN INCA.
OS Chlamydia caviae.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=95302975; PubMed=7783634;
RA Rockey D.D., Heinzen R.A., Hackstadt T.;
RT "Cloning and characterization of a Chlamydia psittaci gene coding for
a protein localized in the inclusion membrane of infected cells.",
RL Mol. Microbiol. 15:617-626(1995).
DR EMBL; L35036; AAC1443.1;
SQ SEQUENCE 355 AA; 38802 MW; 335AAD114D226351 CRC64;

Query Match 100.0%; Score 1733; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.3e-93;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTVSTNDTSPVISRASSPTFGDHGKDFDNKKIIPISIEAPTSSAAAVGAKTAIEPGRSP	60
Db	1	MTVSTNDTSPVISRASSPTFGDHGKDFDNKKIIPISIEAPTSSAAAVGAKTAIEPGRSP	60
Qy	61	LLQRICVLVYIIAIALFVVGIAALVCLYLGVSIVSTPSLILMLAIVLVSFVIVITAIRDG	120
Db	61	LLQRICVLVYIIAIALFVVGIAALVCLYLGVSIVSTPSLILMLAIVLVSFVIVITAIRDG	120
Qy	121	TPSQVVRHMKQIQIQFGENTRLHTAVENLKVNVLSLSEQLNQLHRLSDFGDRLEA	180
Db	121	TPSQVVRHMKQIQIQFGENTRLHTAVENLKVNVLSLSEQLNQLHRLSDFGDRLEA	180

```

RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AF001605; AAD18339.1; -
DR EMBL; AF002217; AAF38399.1; -
DR EMBL; AF002545; BAA98396.1; -
DR HSSP; P05412; 1FOS.
DR TIGR; CP0581; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 390 AA; 42507 MW; F78B8469760A4FD4 CRC64;

Query Match 11.5%; Score 200; DB 16; Length 390;
Best Local Similarity 23.0%; Pred. No. 0.00031;
Matches 87; Conservative 64; Mismatches 142; Indels 86; Gaps 13;

QY 8 TSPVISRASPTGDKGKDFDNKKIIPISTEAPTSSAAVGAKAIEPEGRSPLLQRIY 67
DB 2 SSPVNTTSPAPN-----IPAPTTP-----GIPTT---KPRSSFIEKVI 38

QY 68 LVKTIATAALFVVGIAALVCLYLGSVIS-----TPSL-ILMLAIVSFVIVITARDGT 121
DB 39 ----IVAKYILFAI---AATSGALGTILGSLGALTGGIALLVFFVSMVLLGLILKDSI 92

QY 122 PSQVVRHMKQIQOFGGEN-----TLHTAVENLKAVN----- 154
DB 93 SGGEERLREVSFTSENQRLTVITTTLEFVKDLKAADKDLTLEAFNENGLKTT 152

QY 155 -VELSEQINOLKQLHRLSDFGDRLEANTGDFTLIADFOLSLEEFKS-VGKVTETMLSP 212
DB 153 AEDLEEQSKLSEQLALERINOLIQAAGDAQEISSELKKLISGWDKSVVEQINTSI-- 210

QY 213 FEKLAQSLKETFSQEAQVAMMSVTELTNNAKELI-----TENKTVEIOLKADAQ 265
DB 211 -----QALKVLILGQEWVQEAQTHVKMQEQIQAQAEITLGHMNSTALOKSVENLLVDQ 265

QY 266 LREOVRELEKFOELEECACSTLSHSIATLQESTTLLKDDSTTNLHVESRLIGVMVDGA 325
DB 266 ALTRVVGELLESENKLSQACSAALROEIEKLAQHETSLQORIDAMLAQBNL-----A 317

QY 326 ESSTVFEASQDSDAQPD 344
DB 318 EQVTALEKMKQEAQKAASE 336

RESULT 4
Q9PKR8 PRELIMINARY; PRT; 276 AA.
ID Q9PKR8
AC Q9PKR8;

```

```

RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AF001605; AAD18339.1; -
DR EMBL; AF002217; AAF38399.1; -
DR EMBL; AF002545; BAA98396.1; -
DR HSSP; P05412; 1FOS.
DR TIGR; CP0581; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 390 AA; 42507 MW; F78B8469760A4FD4 CRC64;

Query Match 11.5%; Score 200; DB 16; Length 390;
Best Local Similarity 23.0%; Pred. No. 0.00031;
Matches 87; Conservative 64; Mismatches 142; Indels 86; Gaps 13;

QY 8 TSPVISRASPTGDKGKDFDNKKIIPISTEAPTSSAAVGAKAIEPEGRSPLLQRIY 67
DB 2 SSPVNTTSPAPN-----IPAPTTP-----GIPTT---KPRSSFIEKVI 38

QY 68 LVKTIATAALFVVGIAALVCLYLGSVIS-----TPSL-ILMLAIVSFVIVITARDGT 121
DB 39 ----IVAKYILFAI---AATSGALGTILGSLGALTGGIALLVFFVSMVLLGLILKDSI 92

QY 122 PSQVVRHMKQIQOFGGEN-----TLHTAVENLKAVN----- 154
DB 93 SGGEERLREVSFTSENQRLTVITTTLEFVKDLKAADKDLTLEAFNENGLKTT 152

QY 155 -VELSEQINOLKQLHRLSDFGDRLEANTGDFTLIADFOLSLEEFKS-VGKVTETMLSP 212
DB 153 AEDLEEQSKLSEQLALERINOLIQAAGDAQEISSELKKLISGWDKSVVEQINTSI-- 210

QY 213 FEKLAQSLKETFSQEAQVAMMSVTELTNNAKELI-----TENKTVEIOLKADAQ 265
DB 211 -----QALKVLILGQEWVQEAQTHVKMQEQIQAQAEITLGHMNSTALOKSVENLLVDQ 265

QY 266 LREOVRELEKFOELEECACSTLSHSIATLQESTTLLKDDSTTNLHVESRLIGVMVDGA 325
DB 266 ALTRVVGELLESENKLSQACSAALROEIEKLAQHETSLQORIDAMLAQBNL-----A 317

QY 326 ESSTVFEASQDSDAQPD 344
DB 318 EQVTALEKMKQEAQKAASE 336

RESULT 4
Q9PKR8 PRELIMINARY; PRT; 276 AA.
ID Q9PKR8
AC Q9PKR8;

```


Tue Aug 13 14:23:44 2002

us-09-673-763-8.rspt

```
DR EMBL: AF327003; AAG61100.1; ..
SQ SEQUENCE 273 AA; 30357 MW; 61FA8F2085701F58 CRC64;

Query Match          9.5%; Score 164.5; DB 2; Length 273;
Best Local Similarity 22.8%; Pred. No. 0.023;
Matches 67; Conservative 59; Mismatches 109; Indels 59; Gaps 10;

QY 32 IIPISIEAPTSSAAAVGAKTAIEPEGRSPLLQRICYLVKIIAAIAALFVVGIAALVCLYLIG 91
DQ 8 VTPSPAPPSYANRV-----PQPSLMDKIKKIAAIAASLIIGTIGFLALLHLVG 58
QY 92 SVISTPSLIILMLAIMLVSFVIVITAIRDGTSPQVVR-----HMKQIQOQFGENTRLHTAV 147
DQ 59 FLIAPQITIVLLALFIISLA-----GNALYLQKTANLHLYQDLQ-----EV 100
QY 148 ENKAVNVELSEQINOLKQLHRLSDFGDRLFANTGDTALFADFSLEEFKSVGKTVE 207
DQ 101 GSKLEINFMLS--VLOKEFLH-----LSKEFATTSKDLASVDSQFSCLOQFDRNKGFE 153
QY 208 TMLSPFEKLAQSLKETFSQEAQVAMSSVTELRNLNALKELITENK-----TVI 257
DQ 154 SLDEYKNSTPEMRKLFQSEIIADLKGVSASLREEIRFLTPLAEEVRRLAHNOQSLTVI 213
QY 258 EOLKA-DAQURE--QVRFLK-----RKQLEFACSTLSHSIATLOESTT 300
DQ 214 EELKTRDRLDEIGLSQSLKTLTSQIALQKRESSDLCQSOIRETUSPPKRSAS 267

RESULT 7
ID Q9AM94 PRELIMINARY; PRT; 273 AA.
AC Q9AM94;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9227;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in InCA Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9227;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in InCA Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9301;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in InCA Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=SA-1, APACHE-2, IOL-238, UW-36, 2B, 10A, AND 11A;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Ossewaarde J.M., Morre S.A., van der Brule A.J.C., Dankert J.;
RT "Normal InCA expression and fusogenicity of inclusions in Chlamydia trachomatis isolates with the InCA I47T mutation.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE001286; AAC67710.1; -
DR EMBL: AF279346; AAG31466.1; -
DR EMBL: AF327329; AAK11233.1; -
DR EMBL: AF326992; AAG61089.1; -
DR EMBL: AF326994; AAG61091.1; -
DR EMBL: AF326998; AAG61095.1; -
DR EMBL: AF327002; AAG61099.1; -
DR EMBL: AF327004; AAG61101.1; -
DR EMBL: AF327006; AAG61103.1; -
DR EMBL: AF327012; AAG61109.1; -
DR EMBL: AF327014; AAG61111.1; -
DR EMBL: AF327015; AAG61112.1; -
KW Complete proteome.
SQ SEQUENCE 273 AA; 30327 MW; 6945AE7B8B8BEF3 CRC64;

Query Match          9.5%; Score 164.5; DB 16; Length 273;
Best Local Similarity 22.8%; Pred. No. 0.023;
Matches 67; Conservative 59; Mismatches 109; Indels 59; Gaps 10;

QY 32 IIPISIEAPTSSAAAVGAKTAIEPEGRSPLLQRICYLVKIIAAIAALFVVGIAALVCLYLIG 91
DQ 8 VTPSPAPPSYANRV-----PQPSLMDKIKKIAAIAASLIIGTIGFLALLHLVG 58
QY 92 SVISTPSLIILMLAIMLVSFVIVITAIRDGTSPQVVR-----HMKQIQOQFGENTRLHTAV 147
DQ 59 FLIAPQITIVLLALFIISLA-----GNALYLQKTANLHLYQDLQ-----EV 100
QY 148 ENKAVNVELSEQINOLKQLHRLSDFGDRLFANTGDTALFADFSLEEFKSVGKTVE 207
DQ 101 GSKLEINFMLS--VLOKEFLH-----LSKEFATTSKDLASVDSQFSCLOQFDRNKGFE 153
QY 208 TMLSPFEKLAQSLKETFSQEAQVAMSSVTELRNLNALKELITENK-----TVI 257
DQ 154 SLDEYKNSTPEMRKLFQSEIIADLKGVSASLREEIRFLTPLAEEVRRLAHNOQSLTVI 213
QY 258 EOLKA-DAQURE--QVRFLK-----RKQLEFACSTLSHSIATLOESTT 300
```

RESULT	10	
Q9F7K9		
ID	Q9F7K9	
AC	Q9F7K9	PRELIMINARY;
DT	01-MAR-2001	PRT; 273 AA.
DT	01-MAR-2001	(Tremblrel. 16, Created)
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)
DE	INCLUSION MEMBRANE PROTEIN A.	

RESULT	11	
069196		
ID	069196	PRELIMINARY; PRT; 273 AA.
AC	069196;	
DC	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	INCLUSION MEMBRANE PROTEIN.	
GN	INCA.	
OS	Chlamydia trachomatis.	
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
OX	NCBI_taxid=813;	
	[1]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=LGV-434;	
RC	STRAIN=99043938; PubMed=9826388;	
RA	MEDLINE=99043938; PubMed=9826388;	
RR	Bannantine J.P., Stamm W.E., Suchland R.J., Rockey D.D.;	
RT	"Chlamydia trachomatis INCA is localized to the inclusion membrane and	
RT	is recognized by antisera from infected humans and primates.";	
RL	Infect. Immun. 66:6017-6021(1998).	

QY	32	IIPSTIEAPTSSAAAVGAKTAPEGRSPLLQRICVLYKIIIAAALFVVVGIAALVCYLIG	91
	:	:::::	:
Dd	8	VTPPPAPSYSNRV-----PQPSLMDKKIKKIAIASILGTICTGFTALLHLWG	58
	:	:::::	:
QY	92	SVISTPSLILMLAIAMLVSEVIVITAIRDGPVSQVR---HMKQQIQOFGEEENTRLHTAV	147
	:	:::::	:
Dd	59	FLIAPQITTVLLPALPIISLA-----GNALYLQKTANLHLFDQLQR-----EV	100
	:	:::::	:
QY	148	ENLKAVNVELSEQINOLKQHLTRLSDGDRLEANTGDFTALIADFOLSLEEFKSQVKTYE	207
	:	:::::	:
Dd	101	GSLKEINFMLS-VLQKEFLH----LSKEFATTSKDLSAVSODFYSCLOGFDRDNKGYE	153
	:	:::::	:
QY	208	TMSLPPEKIQAQSLKETFSQFAVQAMMSSVTETRLNALKELETENK-----TVI	257
	:	:::::	:
Dd	154	SLLDEYKNSTEEMKRKFSGEIIDATLDUGSVASIREIRFLPLAEVEVRLAHNQSLTVVI	213
	:	:::::	:

OS	Chlamydia trachomatis.
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX	NCBI_TaxID=813;
RN	[1]
RP	SEQUENCE FROM N. A.
RC	STRAIN=UW-12;
FX	MEDLINE=21295121; PubMed=11402010;
XX	Pannekoek Y., van Der Ende A., Bijl P.P., van Ma

Tue Aug 13 14:23:44 2002

us-09-673-763-8.rspt